A Comparison of Second-Order Classifiers for SAR Sea Ice Discrimination

Abstract

In this paper we present results of an analysis of the relative utility of statistical, structural, and frequency based second-order texture methods for discrimination of sea ice types in synthetic aperture radar (SAR) data. Algorithms were trained using a calibration data set and robustness of the methods were assessed by directly computing ice classes within a validation data set.

Classification using a first-order approach (average grey level) produced Kappa classification accuracies of 51.0 and 33.0 percent for the calibration and validation data. The first-order approach is provided primarily as a reference from which to compare the second-order approaches because the test conditions were selected to be specifically difficult (i.e., different incidence angle ranges between calibration and validation images) for any approach using image tone or the relative scattering cross section.

Results from the second-order approaches indicate that the two spatial domain statistical approaches, Grey Level Co-Occurrence Matrix (GLCM) and the Neighboring Grey Level Dependence Matrix (NGLDM) provided high classification accuracies under the difficult test conditions examined here. The GLCM results achieved a Kappa Coefficient of 84.0 and 81.0 percent for the calibration and validation sets. The NGLDM achieved a Kappa Coefficient of 83.0 and 76.0 percent for the calibration and validation data sets. These results are statistically equivalent between the calibration and validation data sets and between the GLCM and NGLDM schemes. The Spatial/Spatial Frequency (S/SF) approach appears to be sensitive to the training conditions generated from the calibration data set and therefore do not provide statistically reproducible results between the calibration (87 percent) and validation (18 percent) test conditions. Results from the Primitive Texture Value (PTV) method suggest poor operational capabilities both due to low calibration (65 percent) and validation (58 percent) accuracies.

Introduction

There now is a substantial awareness of the economic and environmental significance of the polar regions. Industrial, ecological, and environmental interests have fueled an increase in a variety of polar research disciplines. A major requirement of these activities is timely information on the distribution and dynamics of sea ice. Operational personnel, such as off-shore engineers and icebreaker captains, require knowledge of the location and distribution of sea-ice hazards for route selection and activity scheduling. Ice motion modelers, climatologists, and marine ecologists require knowledge of type, distribution, and seasonal evolution of the sea-ice cover to illuminate processes occurring through the Atmosphere-Cryosphere-Hydrosphere interface.

Monitoring the polar regions on a routine basis at a variety of electromagnetic frequencies is now possible. Current sensors operate at various spatial and temporal resolutions, using energy from microwave, optical, and infrared wavelengths. Active microwave sensing, specifically synthetic aperture radar (SAR), provides the capability for all-weather, day or night monitoring. With the recent launch of the Soviet Almaz, European ERS-1, Japanese JERS-1, and with the impending launch of at least two other SAR satellites (RADARSAT [Canada] and ERS-2 [Europe]),the 1990s are certainly the decade of orbital radar remote sensing.

A critical issue in radar remote sensing science is conceiving a method by which the high data volumes of SAR can be reduced into geophysically meaningful information. In order to process these large volumes of data, there will be a requirement for automated or, at the very least, semi-automated processing. To date, a wide variety of methods have been investigated for machine-assisted classification of SAR data. In general, two approaches to geophysical data extraction are currently being pursued: (1) modeled forward and inverse scattering and (2) statistical classification or retrieval methods. In this work we will deal with the latter approach. Specifically, we will consider two fundamental approaches to statistical classification of SAR images of sea ice. These two approaches are referred to as first- and second-order methods.

First-order approaches make use of image tone, which is a grey-level representation of the relative scattering cross section ($\sigma°$). Statistical separation of sea-ice classes is performed using maximum-likelihood or discriminant clustering types of classification (for example, see Kwok et al. (1992) and Nystuen and Garcia (1992)). Second-order approaches also make use of the grey-level or relative scattering cross sections of a particular SAR resolution cell. The salient difference (relative to first-order approaches) is that a second-order

David G. Barber
Earth Observations Laboratory, Institute for Space and Terrestrial Science, Department of Geography, University of Waterloo, Ontario N2L 3G1, Canada.

Mohammed E. Shokr
Atmospheric Environment Service, 4905 Dufferin Street, Downsview, Ontario M3H 5T4, Canada.

Richard A. Fernandes
Earth Observations Laboratory, Institute for Space and Terrestrial Science, Department of Geography, University of Waterloo, Ontario N2L 3G1, Canada.

Eric D. Soulis
Department of Civil Engineering, University of Waterloo, Ontario N2L 3G1, Canada.

Dean G. Flett
Ellsworth F. LeDrew
Earth Observations Laboratory, Institute for Space and Terrestrial Science, Department of Geography, University of Waterloo, Ontario N2L 3G1, Canada.

approach utilizes a measure of the spatial information which arises from the statistical relationship between each SAR resolution cell and its neighbors. The second-order nature of these statistics means that there is less reliance on the absolute calibration of the image data (i.e., relative rather than absolute calibration is important). A second-order approach will be more sensitive to the spatial variation in the relative scattering cross section as a function of scale.

Research Objectives

One of the major drawbacks of previous statistical classification studies is that validation results were often limited to the data from which the algorithms were calibrated, often a single SAR scene. Comparative analyses are required to assess the robustness of various algorithms within a particular SAR scattering season. Our primary objective in this paper is to evaluate four second-order approaches to image texture representation within a standardized experimental design. Our working hypothesis is that, by evaluating the accuracy and robustness of these texture measures within the winter SAR scattering season (Livingstone et al., 1987), we will be in a position to choose one for implementation within a more comprehensive, operational sea-ice classification system. This evaluation is but a single step in development of an operational sea-ice classification algorithm. Tone, structural information (floe size, shape, morphology, etc.), and data from sequential images are undoubtedly of additional value. The relative importance of these factors will vary but we contend that texture will frequently be useful. We have selected STAR-1 imagery because these data have been extensively validated and because data from a similar system (STAR-2) provides the majority of sea-ice data for the Ice Centre, Environment Canada.

In what follows we review the current status of automated algorithm development, describe the mechanics of the four approaches being contrasted, and conduct a statistical comparison of the relative utility of each for separation of sea-ice types in X-band (9.25 GHz) HH polorized SAR images of sea ice.

Textural Discrimination of SAR Sea-Ice Types

The most common methods of texture analysis involve the use of second-order approaches. Statistical, structural, and frequency based second-order texture methods are used in these analyses and are reviewed here.

Statistical approaches make use of grey-level probability density functions (PDF). The PDF is usually computed as the conditional joint probability of pairs of pixel grey levels in a local area of the image. The Grey Level Co-Occurence Matrix (GLCM; Haralick et al., 1973), Grey Level Run Lengths (Galloway, 1975), and the Neighboring Grey Level Dependence Matrix (NGLDM; Sun and Wee, 1982) are examples of statistical approaches to texture analysis. Structural approaches exploit the notion of a texture primitive, or the basic building blocks of the structural nature of the texture (Gonzalez and Wintz, 1987). These approaches generally consist of finding a texture primitive, extracting a set of features that characterize this primitive, and then defining a placement rule for these primitives within the texture field of the image (Conners and Harlow, 1980b; Shokr and Barber, 1990). Frequency domain analysis is based on the Fourier transform. Analysis is conducted in the frequency domain on the power spectra of the image. This approach is often called PSM (Power Spectrum Method; Conners and Harlow, 1980a). A variety of methods have been developed to separate different textures based on their power spectra (Roan and Aggarwal, 1987; D'Astous and Jernigan, 1984; Matsuyama et al., 1983).

The GLCM has met with some success in a variety of remote sensing applications (Weszka et al., 1976; Shanmugan et al., 1981; Ulaby et al., 1986; Pultz and Brown, 1987; Barber and LeDrew, 1991; Nystuen and Garcia, 1992). Holmes et al. (1984) found that point statistics from the GLCM could provide discrimination of first-year, first-year ridged, multiyear, and multi-year rough sea ice. Classification accuracies, however, were on the order of 50 to 60 percent; lower than what could be considered operationally useful. Shokr (1990) conducted a preliminary investigation into the application of five texture statistics from GLCMs. He found that these texture statistics were more robust to changes in tone between image dates and sensor frequencies when compared with simple tone. Barber et al. (1991) evaluated the robustness of tonal versus GLCM textural classifications of sea ice using data from two coincident images of Mould Bay, N.W.T. This analysis showed that tone was not as robust as texture to variations in grey level created by differences in the look direction of the SAR when imaging the same ice surface.

Recently, researchers have attempted to determine the optimal combination of GLCM parameters for texture computation as well as the optimal texture features or metrics for class separation. Barber and LeDrew (1991) determined that an interpixel sampling distance of one produced significantly better discrimination between ice classes than larger distances of five and nine, and a set of three texture statistics considered simultaneously provided the best discrimination. Shokr (1991) used multi-frequency (L-, C-, and X-band) SAR data to make a first assessment of using universal texture/ice type relationships which hold for all combinations of radar parameters. Five GLCM texture measures were found to be highly correlated and did not vary significantly with the computational variables used in generating the co-occurrence matrices. Combining tone with texture improved the classifications noticeably. Rauste (1990), attempting to improve the separability of thick ice and water or nilas, used 13 measures computed from the GLCM to identify the best feature combination. An increase of 15 percent in classification precision over the original SAR data was observed using the three best texture features. Similar to Barber and LeDrew (1991), no significant improvement in classification precision was observed over the original distance of one pixel. Skriver (1991) used a stepwise discriminant analysis on second-order texture measures computed using the Spatial Grey Level Dependence Method to determine the optimum feature vector from a variety of possible parameters for separating multiyear, smooth first-year, rafted first-year, rough first-year, mixed types, and open water. An improvement of 20 percent was obtained for larger area samples (64 by 64 pixels) when texture measures were included as compared to average intensity alone. Classification of a sub-image using optimum parameters resulted in 16 percent misclassification.

Wackerman (1991) developed a method for optimal linear combination of a given set of image features to achieve maximal separation of classes. Although not designed as a classification algorithm, the method was intended to provide benchmarks for existing algorithms and a means of ranking different sets of feature vectors. Two sets of features—four first-order and six GLCM texture measures—were applied to two different SAR data sets of sea ice. It was found that the texture measure ratios provided the best overall discriminatory capabilities for both data sets.

Other researchers have used a variety of model-based methods to examine texture analysis of SAR sea-ice signatures. Jha and Jernigan (1990) used Markov random fields to model the texture of image regions in SAR images of sea ice.

The model parameters are estimated in the frequency domain by matching the parametric form of the model spectrum to a nonparametric spectrum calculated from the observed data. In addition, speckle was modeled and treated as both a source of noise and information in the segmentation process. Qualitative separation of multi-year, first-year, and new ice was achieved but no quantitative results were given.

A number of previous texture segmentation studies have used linear spatial/spatial frequency (S/SF) representations (Wilson and Spann, 1988; Unser and Eden, 1989; Bovik, 1991) and bilinear representations (Reed and Weschler, 1990). A texture segmentation based on localized spatial filtering formulated in terms of the two-dimensional Gabor function were used by Du (1990) to analyze SAR images of oceans and ice. The functions were chosen to appropriately match the spatial frequency and directional bandwidth of each kind of texture involved. Segmentation was achieved by assigning each pixel to the texture whose filter generated the strongest response at that position. A qualitative comparison of classification results, to those obtained using the GLCM approach, revealed that the co-occurrence method preserved finer texture detail but the Gabor method correctly portrayed the overall locational distributions of ice types/textures and was computationally faster.

Methods

To simplify the classification and thereby exert more control over our experimental condition, we chose first-year smooth (FYS), first-year rough (FYR), and multi-year ice (MYI) as our classification features. Figure 1 depicts examples of the SAR image textures for each of the three ice types extracted from X-band, HH polarization, STAR-1 (Sea Ice and Terrain Assessment Radar-1; Nichols et al., 1986) SAR imagery of Mould Bay, N.W.T. Fifty texture subareas of 625 pixels per sample were extracted for each of the ice types from two SAR scenes (Figure 2a and 2b). All subareas were randomly extracted from 6-m seven-look imagery. The "Calibration" scene was imaged flying north, looking west (Figure 2a; note that the images are mirrored across the east-west vector) and the "Validation" scene was imaged flying west and looking to the south (Figure 2b). Each scene was imaged within 1.5 hours of the other. Grey-tone histograms of the texture subareas are shown in Figure 3.

In the following analysis, robustness is defined as the replicability of a classification when each algorithm is calibrated in a scene with minimal range-dependent tonal variation (calibration image), then directly applied to a second scene (validation image) which incorporates a larger range-dependent variation. The validation image is of the same ice field but imaged at a 90° angle to the calibration image (Figure 2, calibration and validation images). If we assume that each of the texture algorithms is insensitive to monotonic transformations and rotationally invariant, the classification results between calibration and validation images should be statistically equivalent. Note that we have not constrained the classification scheme, only the algorithm used to generate the texture features. The following have been selected for comparative analyses.

Average Grey level (first order approach)

Average grey-tone signatures from the calibration sets were used in a supervised maximum-likelihood classification scheme based on the maximum *a posteriori* probability rule. All calibration samples for a particular ice class (50 by 625 pixels) were used in developing a first-order classification signature. The coefficients from the calibration set were then directly applied to the validation samples.
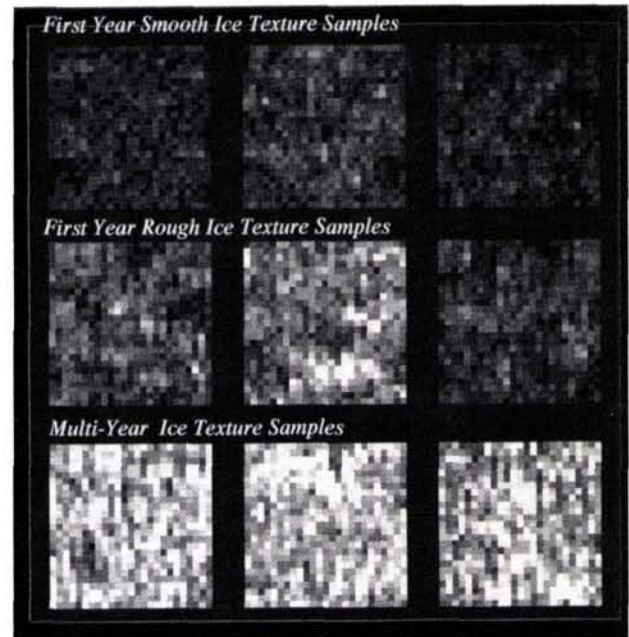

Figure 1. SAR image texture examples: first-year smooth, first-year rough, and multiyear ice. These textures were randomly selected from the 50 calibration samples used in these analyses. Close examination shows the large variability among the three FYR classes, and the relative homogeneity of the FYS and MYI samples.

Neighboring Grey-Level Dependence Matrix (NGLDM)

The Neighboring Grey-Level Dependence Matrix (NGLDM) was proposed by Sun and Wee (1982) as a means to summarize image texture. It is similar to the GLCM but entries have no directional dependence and there are fewer columns in the matrix. The NGLDM, also referred to as the $\mathbf{Q}$ matrix, summarizes the similarity of a neighboring pixel, within a homogeneous class. The relevant neighborhood is defined as an $n \times n$ region centered on the pixel with typical values for n of 3, 4, or 5. A pixel in the neighborhood is defined as similar to the center pixel if its grey level is within a range $\pm \alpha$ grey levels of that of the center pixel. The value of $Q(k,s)$, an element of $\mathbf{Q}$, is the number of pixels with grey level k that have s similar neighbors. $\mathbf{Q}$ has dimensions N_g, the number of grey levels, and S, the maximum number of similar neighbours. S is equal to n_2.

The $\mathbf{Q}$ matrix can be viewed as a two-dimensional extension of the image histogram for a class. When α is very high, all entries in the matrix will be zero except in the n_2 column, which will be the class histogram. As α decreases, these counts will be distributed among the other columns.

Various statistics of $\mathbf{Q}$ summarize the texture of a class and provide features for a classification procedure. Small number emphasis (SNE), given by Equation 1, is

$$
\text{SNE} = \frac{\sum\limits_{k=1}^{N_g} \sum\limits_{s=1}^{S} Q(k,s)/s^2}{\sum\limits_{k=1}^{N_g} \sum\limits_{s=1}^{S} Q(k,s)} \tag{1}
$$

provides a measure of roughness. For a rough texture, the entries in the NGLDM will be primarily in the leftmost column,
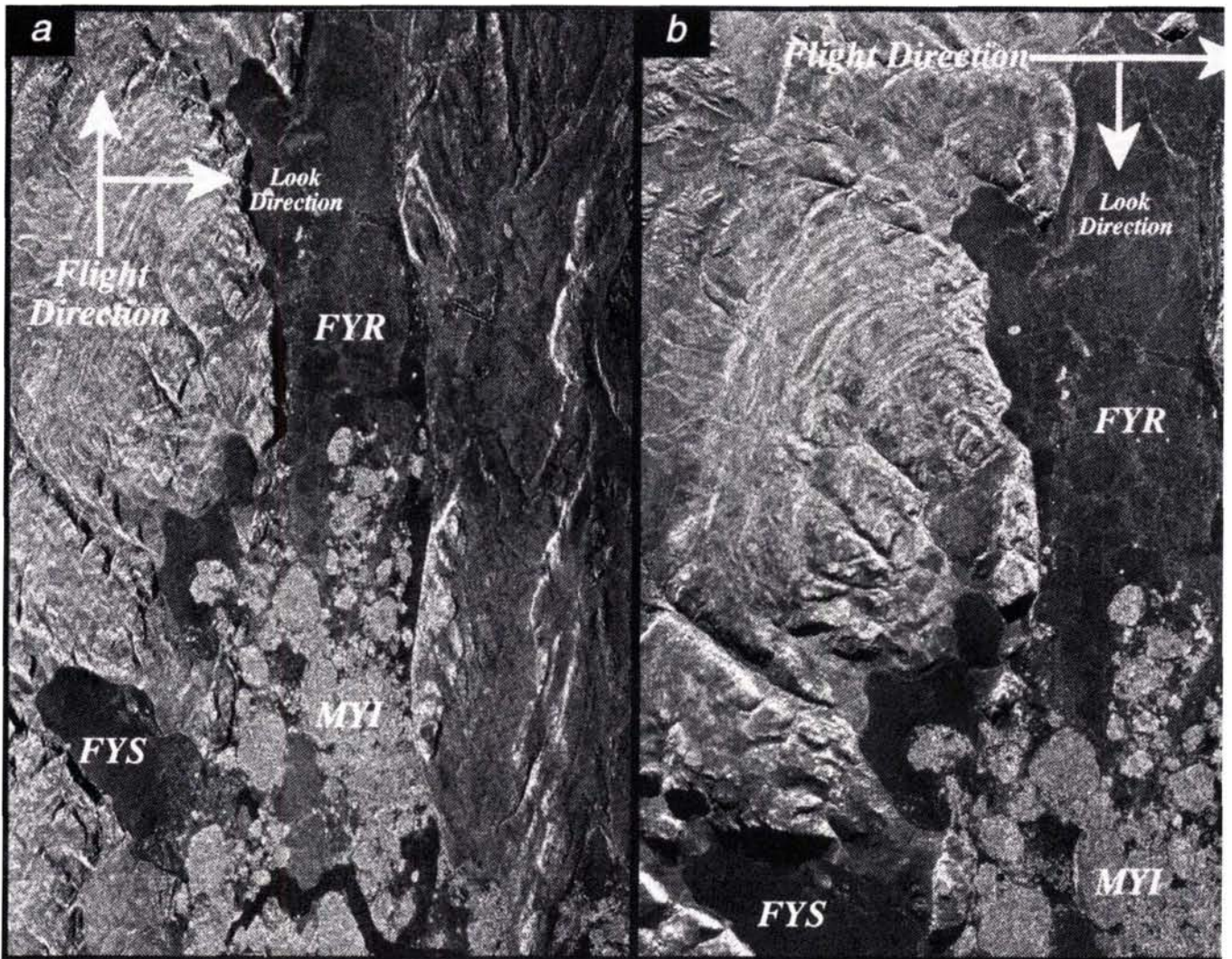
Figure 2. (a) Calibration SAR image. (b) Validation SAR image. Data were acquired with the Sea Ice and Terrain Assessment Radar (STAR-1) over Mould Bay, N.W.T. The images are oriented with north at the top.

and the magnitude of these terms will be emphasized by dividing by s^2. Therefore, larger values of SNE are associated with rougher texture.

Conversely, large number emphasis (LNE), given by Equation 2, is

$$\text{LNE} = \frac{\sum\limits_{k=1}^{N_g} \sum\limits_{s=1}^{S} Q(k,s)s^2}{\sum\limits_{k=1}^{N_g} \sum\limits_{s=1}^{S} Q(k,s)} \qquad (2)$$

provides a measure of smoothness. As a result, smoother image texture produces larger LNE.

Number non uniformity (NNU), second moment (SMT), and entropy (ENT), given by Equations 3, 4, and 5, provide measures of homogeneity. That is,

$$\text{NNU} = \frac{\sum\limits_{s=1}^{S} \left[\sum\limits_{k=1}^{N_g} Q(k,s) \right]^2}{\sum\limits_{k=1}^{N_g} \sum\limits_{s=1}^{S} Q(k,s)} \qquad (3)$$

$$\text{SMT} = \frac{\sum\limits_{k=1}^{N_g} \sum\limits_{s=1}^{S} Q(k,s)^2}{\sum\limits_{k=1}^{N_g} \sum\limits_{s=1}^{S} Q(k,s)} \qquad (4)$$

$$\text{ENT} = \frac{\sum\limits_{k=1}^{N_g} \sum\limits_{s=1}^{S} Q(k,s) \log(Q(k,s))}{\sum\limits_{k=1}^{N_g} \sum\limits_{s=1}^{S} Q(k,s)} \qquad (5)$$

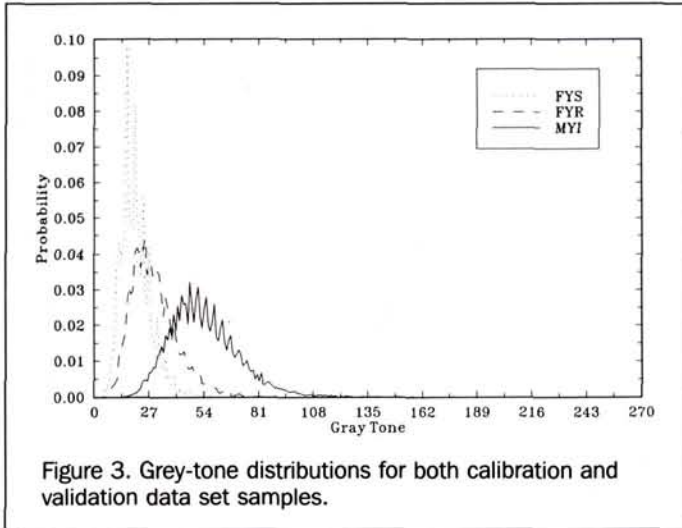The calibration data were used to select the parameter α and

Figure 3. Grey-tone distributions for both calibration and validation data set samples.

to examine the effectiveness of texture statistics of various features and feature pairs for discrimination of the training samples. Alternatives were compared using the results of a leave-one-out, seven-nearest-neighbor classification scheme. This non-parametric approach requires no assumptions regarding the probability distributions or the covariance structures of the features. All single-feature and two-feature combinations were tested with α ranging from 0 to 50 grey levels (the 90 percent grey-level range of the data is 70). A 3 by 3 neighborhood was used based on the experience of McKillop *et al.* (1990) with the same imagery. The set of features for the calibration data set that generated the best Kappa score over a wide range of α was used to classify the validation data, also using a seven-nearest-neighbor scheme.

Primitive Texture Values (PTV)

The primitive texture value (PTV) method is based on a concept for texture quantification introduced by Haralick (1979). The concept entails that texture be defined in terms of two aspects: a primitive element which is described by a tonal value, and the spatial distribution of the primitive tonal values within a neighborhood.

The primitive element is defined as a square of four connected pixels. Hence, six possible pixel-pair combinations are available. The absolute difference between grey tones of pixels in each pair combination is calculated. This is denoted by D_i, where $i = 1, 2, ..., n$; in this case, $n = 6$. Let r denote the upper limit of D_i which is 255 in this application. A numeric string T can then be defined as the set of D_i values, arranged in descending order: i.e.,

$$T = \{D_i/D_{i+1} < D_i; i = 1, 2,..., n\} \qquad (6)$$

where $0 < D_i < r$.

It can be seen that both D_i and T are invariant as a result of adding an offset to the grey tone, but they vary as a result of multiplying the grey tone by a constant. A mapping scheme, S, is then used to map T into a single integer value, called the Primitive Texture Value (PTV): i.e.,

$$S : T \rightarrow \text{PTV}; \qquad 1 < \text{PTV} < L \qquad (7)$$

The scheme is based on a model of placing r objects randomly and distinguishably into n cells, with replacement. The rationale is explained by analogy. Suppose that there are r distinct objects, each of which represents, and are labeled

with, one possible value of D_i. Also, suppose that there are n cells, each of which can accommodate only one object. The construction of T can therefore be envisioned as a process of placing one of the r objects into each cell, until all the cells are filled. To account for the fact that the order of D_i in T is insignificant, only distinguishable placements have to be considered. To account for the fact that within T, a certain D_i value may be repeated, the placement process has to be performed "with replacement." The upper limit of the PTV, defined as L in the above equation, represents the total number of all possible distinguishable combinations of D_i in T. Therefore, it can be obtained using the following combinatorial expression (Feller, 1968):

$$L = (n + r - 1)^C(r - 1) \qquad (8)$$

This expression was used to devise the PTV mapping scheme. The idea is to assign an integer value to each individual component D_i in T, and then add all those integers to form the PTV. The integer assignment is based on the above combinatorial expression, where r and n are replaced by D_i and $(n - i + 1)$, respectively. The mapping equation can therefore be written as

$$\text{PTV} = 1 + \sum_{i=1}^{n} (n + D_i - 1)^C(D_i - 1) \qquad (9)$$

The PTV measure is an indicator of the grey-tone contrast within a primitive element. Texture-rich regions of an image will typically have large PTVs and vice versa. It is insensitive to the order of grey tone within the primitive (see Equation 6), and therefore cannot be used to resolve grey-tone directionality. The assumption implied in this technique is that the primitive, as an atomic element, retains no information on texture directionality. The latter can be addressed when examining the spatial distribution of the PTV. In order to reduce the computation time, the number of components in T was reduced from six to three. This was performed at three stages ($k = 1, 2,$ and 3) using the following recurrence relation:

$$D_i^k = (D_i^{k-1} + D_{i+1}^{k-1})/2 \text{ for } 1 < k < n. \qquad (10)$$

For each sample, the PTV was calculated at each pixel using a 2 by 2 primitive element as described above. The element was selected such that the reference pixel is at the upper-left corner. The average of the PTVs over the sample size (25 by 25 pixels) was then computed and assigned to the sample. Samples were represented in two-dimensional feature space encompassing the PTV and the grey tone to perform the classification. Further details are available in Shokr and Barber (1990).

Spatial/Spatial Frequency (S/SF)

This method uses a linear Spatial/Spatial Frequency (S/SF) representation, generated by Gabor filters, for texture discrimination. Wright (1988) showed that a suitably parameterized basis of two-dimensional Gabor functions can be weighted to create images that are visually indiscriminable from certain Brodatz textures. Gabor filters have maximally smooth windows in the frequency domain. This property prevents irregular distortions of frequency components. These points suggest that a family of Gabor filters may offer a useful characterization of the S/SF energy in an image.

A Gabor filter is specified in the frequency domain as

$$G(r,\theta) = \frac{1}{\sigma r \sigma \theta \sqrt{2\pi}} \exp\left[\frac{(r - r_0)^2}{\sigma_r^2} + \frac{(\theta - \theta_0)^2}{\sigma_\theta^2}\right] \qquad (11)$$

where

r = radial frequency in cycles per degree (cpd),
r_0 = radial center frequency in cpd,
σ_r = radial bandwidth in octaves,
θ = orientation in radians,
θ_0 = center orientation in radians, and
σ_θ = orientation bandwidth in radians.

Wright's parameterization makes use of 42 filters at seven orientations:

$$\theta = \left[\frac{7\pi}{14}, \frac{5\pi}{14}, \frac{3\pi}{14}, \frac{\pi}{14}, \frac{-3\pi}{14}, \frac{-5\pi}{14}\right] \text{ radians} \quad (12)$$

with $\sigma_\theta = \pi / 14$; and six radial frequencies

$$r = [1.7, 2.8, 4.3, 7.1, 10.3, 13.0] \text{ cpd} \quad (13)$$

with corresponding σ_r's:

$$\sigma_r = [\ 2.0, 1.9, 1.7, 1.6, 1.5, 1.3] \text{ octaves} \quad (14)$$

The 13.0 cpd filters were discarded due to insufficient resolution. The field of view for a 32 by 32 pixel mask was set at 1.5534 degrees, so that the Nyquist frequency was exactly 10.3 cpd—the highest center frequency of the filters used. This setting gives almost no aliasing from the 10.3 cpd filters when their energy outputs are halved. Finally, the energy from each filter is normalized by its area within spatial frequencies 0 and 10.3 cpd.

Each sample was DC normalized and the magnitude of its two-dimensional FFT was computed. The resulting spectrograms were scaled by each Gabor mask and the output summed to give the 35 Gabor energy measurements (E_1, ..., E_{35}). D'Astous (1983) describes a number of S/SF features which may be useful for texture discrimination. These features, adapted to make use of the Gabor filters, are entropy of the vector of Gabor energies, the eigenvalues of the spread of Gabor energy, the sum of energy output from all filters at a given r_0, and the r_0 of the Gabor filter with highest energy output. The features from samples of both data sets were normalized by the maximum absolute value across all samples in the training set. This prevented the occurrence of numerical errors during classification due to disparate ranges in inter-features magnitudes.

Features with inter- to intra-class scatter ratios more than two orders of magnitude below the largest scatter ratio were discarded. Feature pairs with Pearson correlation coefficients greater than 0.9, within class, were grouped together. The features in each group were replaced with the feature with the greatest scatter ratio within the group. This resulted in the two features, F1 and F2:

F1, the largest eigenvalue of the spectral spread matrix; the spread matrix is defined as

$$\sum_G = \begin{bmatrix} \sigma_{20} & \sigma_{11} \\ \sigma_{11} & \sigma_{02} \end{bmatrix} \quad (15)$$

where

$$\sigma_{ij}^2 = \sum_u \sum_v /u - u_o/^i/v - v_o/^i G(u,v); \quad (16)$$

and F2, the sum of energy output from 10.3 cpd filters.

The Wilk W-statistic (Shapiro and Gross, 1981) was used to test the hypothesis that each feature class-conditional distributions were Gaussian. Gaussian maximum-likelihood classification (MLC) was performed with both features and then with only one feature in the event that the features were

not bivariate normal. A pairwise optimal linear discriminant classifier (OLD) was also designed for non-parametric classification using both features (Fukunaga, 1990). The calibration data were classified using the leave-one-out method for MLCs. All other classifications were done with a classifier designed using all calibration samples.

Grey-Level Co-Occurrence Matrix (GLCM)

A Grey-Level Co-Occurrence Matrix (GLCM) provides the conditional joint probabilities of all pairwise combinations of pixels within a computation window (W_{nm}). The co-occurrence of grey levels represents the probability of any two pairs of grey levels occurring, conditional on the interpixel sampling distance (δ) and orientation (α) used for computation. Algebraically, this can be expressed as Equation 17 where C_{ij} is defined in Equation 18 following Haralick et al. (1973): that is,

$$Pr(x) = \{C_{ij}|\delta, \alpha\} \quad \text{and} \quad (17)$$

$$C_{ij} = \frac{P_{ij}}{\sum_{ij=1}^{n} P_{ij}} \quad (18)$$

where P is the frequency of occurrence of grey levels i and j. Summation over n refers to the total number of pixel pairs, which is dependent on the parameter d and the subscripts n and m from W_{nm}.

The texture statistics represent a single measure of the image texture from which the GLCM is computed. Each texture statistic is considered to be a point estimate because it provides a single measure of the total information content of the GLC Matrix. Three texture statistics used in this research are given by Equations 19 to 21.

Uniformity (UNF)

$$\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}^2 \quad (19)$$

When C_{ij}s are large, Uniformity is large. The joint probabilities will be larger when there is less local variation in the texture. This measure is sensitive to the level of homogeneity within the texture field.

Entropy (ENT)

$$-\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \log C_{ij} \quad (20)$$

Entropy is a measure of the classical information theory approach developed by Claude Shannon (Thomas, 1981). Entropy can be rationalized as a measure of disorder. As the pattern in the input image approaches a truly random state, the entropy measure approaches maximum. For homogeneous or simple patterned data, entropy is very small.

Dissimilarity (DIS)

$$\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}/(i - j)/ \quad (21)$$

The more dissimilar the grey levels are at δ and α, the larger this statistic will be. This measure is sensitive to both grey-level spatial variability and tone of the input image.

GLCM texture statistics were evaluated for two- and three-variable cases at three values of α (0°, 45°, and 90°) and δ (1, 3, and 5). Probability plots of the individual GLCM sta-

tistics, within each ice type, were computed. Each was found to be unimodal and statistically indistinguishable from a Gaussian distributed variable. Uniformity, Entropy, and Dissimilarity are pairwise correlated to differing degrees over the various combinations of feature and ice class variables. Minimum pairwise Pearson correlation coefficients were observed at 0.88 to a maximum of 0.96. Results from the best pairwise and three-variable case are provided here. Consistent with previous analysis (Barber *et al.*, 1991; Barber and LeDrew, 1991) results indicated that δ of 1 and an α parallel to the look direction of the radar optimized precision of the classification features. Classifications were computed using Linear Discriminant Analysis (LDA). Rationale and details of the approach are given in Barber and LeDrew (1991).

Results and Discussion

Classification results are presented as contingency tables for each method. The Kappa Coefficient (Cohen, 1960; Rosenfield and Fitzpatrick-Lins, 1986; Story and Congalton, 1986; Congalton, 1991) is used as the statistical metric for intercomparison of algorithms across the calibration and validation data sets.

Average Grey Level (First-Order Approach)

As expected, the classification results from the first-order approach were poor. Kappa classification accuracies were 51 and 33 percent for the calibration and validation data sets, respectively (Tables 1 and 2). The difference between calibration and validation sets is statistically significant at the 95 percent level against a Type I error. Errors in classification are caused mainly by the large variability of tonal signatures of the first-year rough class (Figure 3). For example, (1) surface roughness from compressional forces acting on the first-year ice create high scattering returns, very similar in magnitude to those arising from the volume scattering within hummocks of multiyear ice, and (2) smooth portions, within the rough first-year ice classes, overlap substantially with the first-year smooth case.

Neighboring Grey-Level Dependence Matrix (NGLDM)

The classification statistics of the calibration data set were comparable for all single-feature and two-feature cases for wide ranges of α, the similarity range. The best classification accuracies for each combination varied from 87 to 90 percent. The large number emphasis (LNE) showed the strongest relationship between calibration and validation results, possibly because smoothness is a class characteristic that is least affected by changes in the dynamic range between the two data sets.

LNE had the best precision for α equal to seven grey levels, which is 10 percent of the 90 percent grey-level range of the calibration data. The classification precision for the calibration and validation data sets were 83.0 and 76.0 percent, respectively. The validation results were within the 95 percent confidence intervals for the calibration results. Histograms of LNE for the two sets of data are shown in Figures 4 and 5. Contingency table results are shown in Tables 3 and 4.

Primitive Texture Values (PTV)

A supervised classification algorithm based on the Maximum *a posteriori* Probability rule (Duda and Hart, 1973) was applied to classify a given sample. Statistical signatures (mean, standard deviation, and correlation matrix) of both grey tone and PTV were derived from the calibration data for

TABLE 1. CONTINGENCY TABLE FOR GREY TONE: CALIBRATION DATA

	FYS	FYR	MYI	Total
FYS	43	7	0	50
FYR	25	19	6	50
MYI	1	10	39	50
Total	69	36	45	150

TABLE 2. CONTINGENCY TABLE FOR GREY TONE: VALIDATION DATA

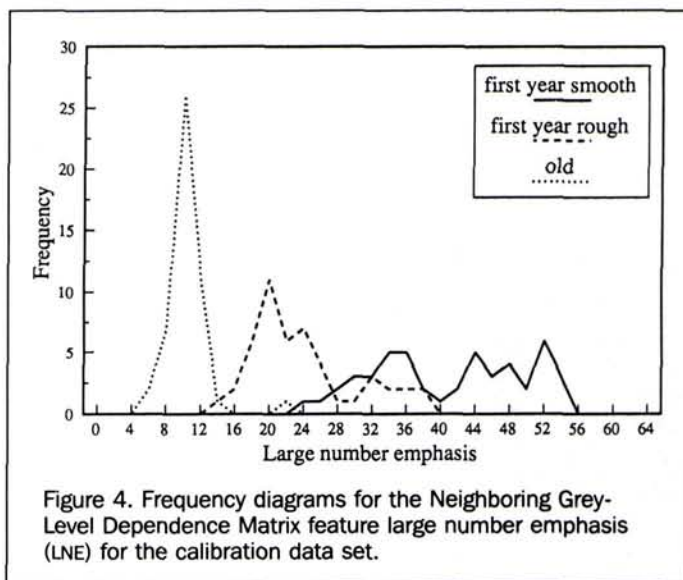|	FYS	FYR	MYI	Total
FYS	41	8	1	50
FYR	27	21	2	50
MYI	6	23	21	50
Total	74	52	24	150

Figure 4. Frequency diagrams for the Neighboring Grey-Level Dependence Matrix feature large number emphasis (LNE) for the calibration data set.
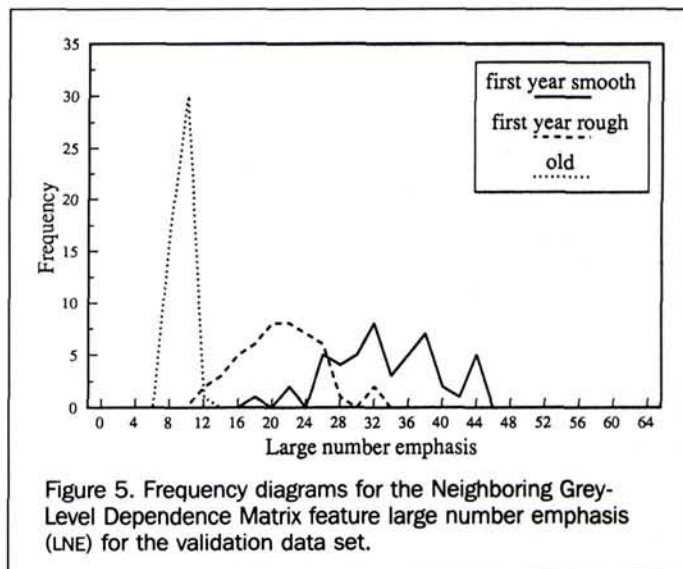

Figure 5. Frequency diagrams for the Neighboring Grey-Level Dependence Matrix feature large number emphasis (LNE) for the validation data set.

TABLE 3. CONTINGENCY TABLE FOR NGLDM LARGE NUMBER EMPHASIS (LNE): CALIBRATION DATA.

	FYS	FYR	MYI	Total
FYS	45	5	0	50
FYR	11	39	0	50
MYI	0	1	49	50
Total	56	45	49	150

TABLE 4. CONTINGENCY TABLE FOR NGLDM LARGE NUMBER EMPHASIS (LNE): VALIDATION DATA.

	FYS	FYR	MYI	Total
FYS	36	14	0	50
FYR	2	40	8	50
MYI	0	0	50	50
Total	38	54	58	150

TABLE 5. CONTINGENCY TABLE FOR PTV: CALIBRATION DATA.

	FYS	FYR	MYI	Total
FYS	49	1	0	50
FYR	21	26	3	50
MYI	0	10	40	50
Total	70	37	43	150

TABLE 6. CONTINGENCY TABLE FOR PTV: VALIDATION DATA.

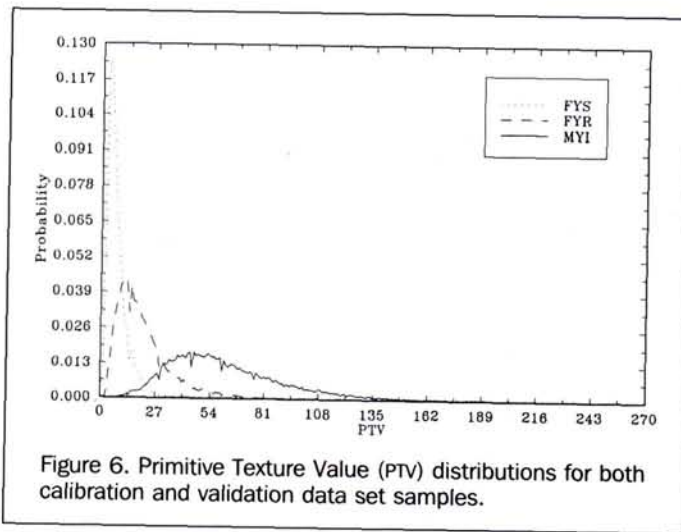|	FYS	FYR	MYI	Total
FYS	47	3	0	50
FYR	22	25	3	50
MYI	1	13	36	50
Total	70	41	39	150

Figure 6. Primitive Texture Value (PTV) distributions for both calibration and validation data set samples.

each class. Assuming that the *a priori* probabilities of individual sea-ice classes are equal, and that the probability density function for the PTV conditioned on an ice class is Gaussian, then the *a posteriori* probabilities can be calculated according to Bayes Rule using the statistical signatures

of ice classes. Each individual sample was assigned to the class associated with the maximum *a posteriori* probability.

Results show that classification accuracy based on PTVs was generally low for the calibration set (65 percent) and even lower for the validation set (58 percent). Tables 5 and 6 show the contingency table results for both the calibration and validation data sets. Misclassifications were high both between FYS and FYR and between MYI and FYR. By far the larger of the two was the separation of the two first-year ice types (FYS and FYR).

Figure 6 illustrates the probability histograms of the PTV for the three ice classes. The histograms were derived from all pixel data in both the calibration and validation samples. Compared to grey tone (Figure 3), the PTV produced smoother histograms for all classes, with noticeable smoothing in the FYS ice histogram. The MYI ice PTV histogram had a high variance and the FYS ice histogram had a somewhat lower variance. Histograms are skewed towards the upper end of each distribution (particularly the MYI distribution). This suggests that the PTV is more sensitive to texture variation at the high (rough) end of the texture scale. This explains the relatively small variance of the FYS ice data compared to the MYI data. The PTV is therefore more successful in distinguishing FYS ice from the other two types (Table 5). The larger variance of the other two types, however, deteriorates the overall performance of the PTV algorithm.

Spatial/Spatial Frequency (S/SF)
The MLC gave optimal results using only the F2 texture feature. A scatter plot of calibration data in the F1-F2 space is shown in Figure 7. It suggests that the class conditional distributions are not bivariate normal. This may explain why the MLC was more precise with only F2. The OLD first assigned samples to MYI because this class had the largest average pairwise scatter ratio. The remaining samples were assigned to FYR or FYS.

Tables 7 to 10 give contingency tables which describe the results for the single feature MLC and the OLD. The OLD was more precise than the MLC in classifying calibration data. This is due to the non-parametric nature of OLD and its use of re-substitution. Both classifiers were significantly less
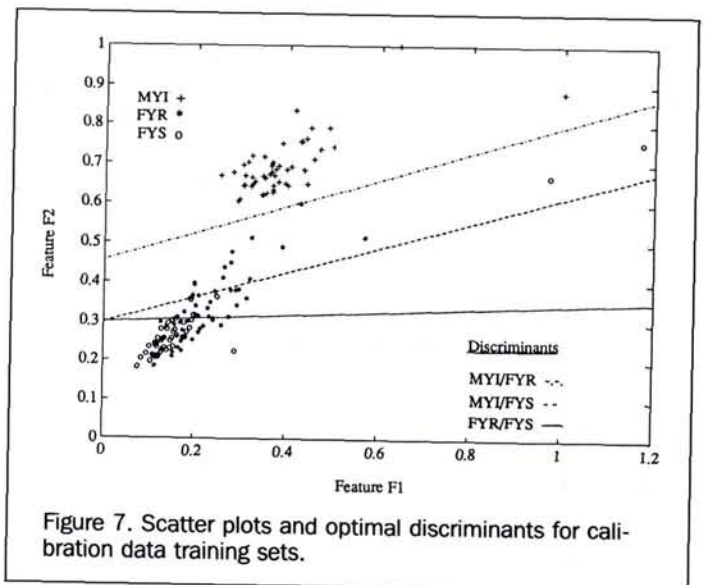

Figure 7. Scatter plots and optimal discriminants for calibration data training sets.

TABLE 7. CONTINGENCY TABLE FOR THE S/SF ML CLASSIFIER, F2 FEATURE: CALIBRATION DATA.

	FYS	FYR	MYI	Total
FYS	38	10	2	50
FYR	2	46	2	50
MYI	0	0	50	50
Total	40	56	54	150

TABLE 8. CONTINGENCY TABLE FOR THE S/SF ML CLASSIFIER, F2 FEATURE: VALIDATION DATA.

	FYS	FYR	MYI	Total
FYS	16	19	15	50
FYR	12	36	2	50
MYI	11	11	28	50
Total	39	66	45	150

TABLE 9. CONTINGENCY TABLE FOR THE S/SF OLD CLASSIFIER: CALIBRATION DATA.

	FYS	FYR	MYI	Total
FYS	45	5	0	50
FYR	8	42	0	50
MYI	0	0	50	50
Total	53	47	50	150

TABLE 10. CONTINGENCY TABLE FOR THE OLD CLASSIFIER: VALIDATION DATA.

	FYS	FYR	MYI	Total
FYS	18	11	21	50
FYR	16	32	2	50
MYI	14	18	18	50
Total	48	61	41	150
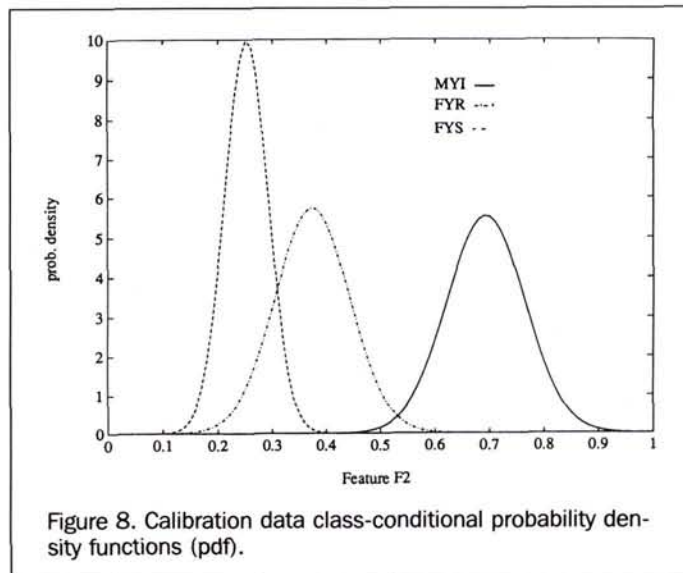
Figure 8. Calibration data class-conditional probability density functions (pdf).

precise when classifying validation data in comparison to calibration data.

Tests of the hypotheses that the means and variances of the F2 class-conditional distributions were equal between calibration and validation were rejected at greater than 95 percent significance levels in all but one instance (a result of the range dependence of these data). This indicates that the drop in precision between calibration and validation data sets was probably due to changes in feature values and not artifacts of the classification methods.

The drop in precision, observed in these results, may be explained by considering a calibration sample consisting of only a sine wave grating. This grating will undergo a non-linear stretch along the look direction in the validation image. This changes the S/SF content of the sample. It may be that the features used here are sensitive enough to respond to these changes in S/SF energy. Feature F2 indicated that the high frequency energy of the ice type samples increases from the FYS to FYR to MYI classes. This is evident in the radar image of Figure 2b. The stretching induced in the validation image will reduce the high frequency energy in all samples. This may explain the increase in MYI misclassifications for validation data.

The variance of class-conditional distributions for both features was more than three times larger in the validation data. Figure 8 indicates that the variance of the FYS class was much lower than that of FYR and MYI. This may explain the tendency of both classifiers to misclassify FYS validation samples. Finally, an increase in incidence angle range within a sample increases F2 because it is an energy spread measure. This may explain the jump in FYS validation sample misclassification from the MLC to the OLD classification schemes.

Grey-Level Co-Occurrence Matrix (GLCM)

The classification statistics of the calibration and validation data were comparable for all pairwise and three-way feature sets tested. The classification accuracies for each combination varied from 80 to 90 percent. Kappa Coefficients of Uniformity and Entropy provided the largest pairwise classification score for both calibration and validation sets (84 and 81 percent, respectively). A three-way discrimination provided marginally poorer results with a calibration precision of 84 percent and a validation Kappa coefficient of 74 percent. This is in contrast to previous work, conducted using only a single SAR scene (Barber and LeDrew, 1991), where a three-variable set provided better discrimination. Misclassifications were generally larger between FYR and FYS in the two-feature space discrimination. Both FYR versus FYS and MYI versus FYR were areas of significant misclassifications in the three variable discrimination. In general, the larger the calibration precision, the smaller the corresponding validation result.

A classification plot, showing the 90 percent confidence ellipses of the bivariate distributions of Uniformity and Entropy, illustrates the separation of each ice class in this feature space (Figure 9). The elliptical shape of the bivariate confidence intervals also reflects the pairwise correlation structure of these variables. Contingency tables for the two-feature and three-feature calibration and validation results are shown in Tables 11 to 14.

Comparative Analysis

A comparison of the Kappa Coefficients and their associated confidence intervals show that the two spatial domain statistical texture approaches—the Grey-Level Co-Occurrence Ma-
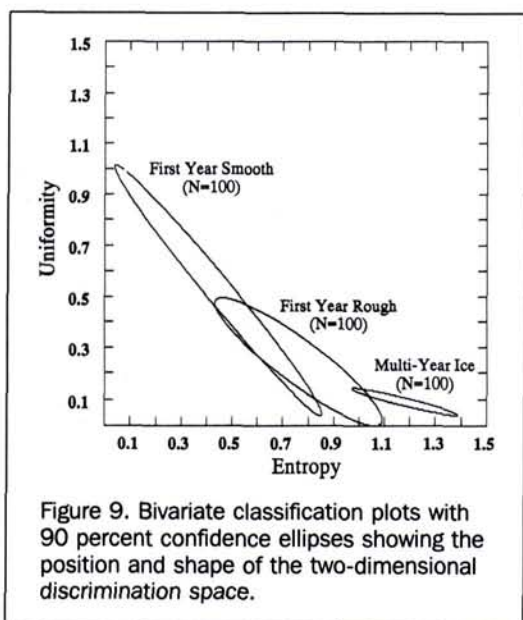
Figure 9. Bivariate classification plots with 90 percent confidence ellipses showing the position and shape of the two-dimensional discrimination space.

TABLE 11. CONTINGENCY TABLE FOR THE GLCM UNIFORMITY AND ENTROPY: CALIBRATION DATA.

	FYS	FYR	MYI	Total
FYS	40	10	0	50
FYR	3	44	3	50
MYI	0	0	50	50
Total	43	54	53	150

TABLE 12. CONTINGENCY TABLE FOR THE GLCM UNIFORMITY AND ENTROPY: VALIDATION DATA.

	FYS	FYR	MYI	Total
FYS	48	2	0	50
FYR	13	37	0	50
MYI	0	4	46	50
Total	61	43	46	150

TABLE 13. CONTINGENCY TABLE FOR THE GLCM UNIFORMITY-ENTROPY-DISSIMILARITY: CALIBRATION DATA.

	FYS	FYR	MYI	Total
FYS	41	9	0	50
FYR	5	43	2	50
MYI	0	0	50	50
Total	46	52	52	150

TABLE 14. CONTINGENCY TABLE FOR THE GLCM UNIFORMITY-ENTROPY-DISSIMILARITY: VALIDATION DATA.

	FYS	FYR	MYI	Total
FYS	49	1	0	50
FYR	15	35	0	50
MYI	0	10	40	50
Total	64	46	40	150

trix (GLCM) and Neighboring Grey-Level Dependence Matrix (NGLDM)—provided the best results within the constraints imposed by this analysis. The GLCM achieved a Kappa Coefficient of 84 and 81 percent for the calibration and validation data sets, respectively. Based on an assumed Gaussian distribution of the Kappa point estimate, these two results are statistically equivalent at the 95 percent level against a Type I error (Table 15). Note that the confidence intervals are computed for the validation, as opposed to the calibration, data. The NGLDM results achieved a Kappa Coefficient of 83 and 76 percent for the calibration and validation sets, respectively. These results are statistically equivalent over the calibration and validation sets within the NGLDM comparison data. They are also statistically equivalent to the GLCM results (Table 15).

The Spatial/Spatial Frequency (S/SF) results achieved good calibration data set results—84 and 87 percent for the F2 and OLD results—but fared poorly when these algorithms were applied to the validation data set (i.e., 30 and 18 percent for F2 and OLD results with the validation data; Table 15). The Primitive Texture Value (PTV) results were generally low for the calibration (65 percent) and validation sets (58 percent). Features F2 and OLD from the S/SF algorithm were statistically different between calibration and validation data. The PTV results were not statistically separable between calibration and validation sets.

In the particular conditions of this test, we have minimized the range of information available for "training" the various algorithms used here. The validation set contains a significantly greater range and variability of both first- and second-order statistical distributions. All three classes were tested for the hypothesis that the variance of the calibration and validation sets were equal. This hypothesis was rejected at the 99 percent significance level, for all three classes, using a likelihood ratio test suggested by Srivastava (1983). A subsequent test for the hypothesis that the means of each class were identical for training and validation sets was also carried out (Srivastava, 1983). The means for each of the FYS, FYR, and MYI violated this hypothesis at the 90 percent level.

Conclusions and Recommendations

These results indicate that the two spatial domain statistical methods, the Grey-Level Co-Occurrence Matrix (GLCM) and the Neighboring Grey-Level Dependence Matrix (NGLDM), provided universally high classification accuracies under the test conditions examined here. The GLCM scores of 84 and 81 percent and the NGLDM scores of 83 and 76 percent were statistically indistinguishable across algorithm and data set factors.

The empirical probability density functions generated with the two spatial domain statistical approaches appear to be reasonably robust to the conditions tested here. It is inter-

TABLE 15. KAPPA COEFFICIENT COMPARISON OF ALGORITHMS.

Algorithm	K_{cal}	$K_{\sigma cal}$	K_{val}	$K_{\sigma val}$	Variable Set	95 percent C.I.
Grey Level	51	0.30	33	0.38	First-Order	39.9–88.9
GLCM	84	0.14	81	0.16	*Unif.-Ent.	73.1–88.9
NGLDM	83	0.15	76	0.20	LNE	67.3–84.7
GLCM	84	0.14	74	0.21	Unif.-Ent-Diss.	64.9–83.0
PTV	65	0.27	58	0.23	PTV	55.0–74.9
S/SF	84	0.14	30	0.37	F1	55.0–74.9
S/SF	87	0.12	18	0.37	F1-F2	6.1–29.9

* Unif = Uniformity; Ent = Entropy; Diss = Dissimilarity.

esting to note that both GLCM and NGLDM approaches resulted in very similar calibration and validation results. This indicates that both algorithms are relatively invariant to monotonic transformations. We also speculate that either these approaches are insensitive to changes in the second-order statistics of the texture fields, or that there simply was no change in the spatial statistics of the image texture.

The Spatial/Spatial Frequency (S/SF) approach appears to be sensitive to the training conditions generated from the calibration data set and therefore does not provide statistically reproducible results between the calibration (87 percent) and validation (18 percent) test conditions. The Primitive Texture Value (PTV) appears to be sensitive to specification of either the texture primitive or the spatial arrangement of this variable within the ice class of interest. Results from the PTV suggest poor operational capabilities due both to low calibration (65 percent) and to validation (58 percent) results.

These results indicate that spatial statistical methods hold promise as a machine directed or machine assisted method for discrimination of ice types in SAR images of sea ice. Our intention here has been to establish the ordination of these four techniques by standardizing a test scenario and focusing on a simple subset of the classification problem, thereby allowing a meaningful statistical evaluation. We have not considered the complexities of the seasonal evolution of SAR signatures, the calibration of aerial and orbital radar data (both absolute and relative), or the extension of the classification problem beyond simple, World Meteorological Organization (WMO) ice-type discrimination schemes.

From our experiences gained through visual interpretation of SAR sea-ice images, automated classification research, and *in situ* validation of scattering signatures, it has become apparent that a multi-level classification scheme is required. We feel an optimal solution for an operational ice classification algorithm from SAR data should consist of a hybrid system (for example, see Wells and Haygen (1991)). An automated, semi-automated, and expert system or knowledge-based classification scheme would provide the three levels required to meet the complexities of the perceived classification problem. The results presented here are a contribution to the automated component of such a system.

Acknowledgments

References

Barber, D.G., and E.F. LeDrew, 1991. SAR Sea Ice Discrimination Using Texture Statistics: A Multivariate Approach. *Photogrammetric Engineering & Remote Sensing*, 57(4):385–395.

Barber, D.G., D.D. Johnson, and E.F. LeDrew, 1991. Measuring Climatic State Variables from SAR Images of Sea Ice: The SIMS SAR Validation Site in Lancaster Sound. *Arctic*, 44(Supp. 1): 108–121.

Bovik, A.C., 1991. Analysis of Multichannel Narrow-Band Filters for Image Texture Segmentation. *IEEE Transactions on Signal Processing*, 39(9):2025–2043.

Cohen, J., 1960. A Coefficient of Agreement of Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

Congalton, R.G., 1991. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of the Environment*. 37:35–46.

Conners, R.W., and C.A Harlow, 1980a. A Theoretical Comparison of Texture Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(3):204–222.

———, 1980b. Toward a Structural Textural Analyzer Based on Statistical Methods. *Computer Graphics and Image Processing*, 12:224–256.

D'Astous, F., 1983. *Textural Feature Extraction in the Spatial Frequency Domain*. Unpublished Ph.D. Thesis, Department of Systems Design, University of Waterloo.

D'Astous, F., and M.E. Jernigan, 1984. Texture Discrimination based on Detailed Measures of the Power Spectrum. *Seventh International Conference on Pattern Recognition*, pp. 83–86.

Du, Li-jen, 1990. Texture Segmentation of SAR Images Using Localized Spatial Filtering. *IGARSS'90*, 3:1983–1986.

Duda, R.O., and P.E. Hart, 1973. Pattern classification and scene analysis. John Wiley & Sons, Inc., New York.

Feller, W., 1968. *An Introduction to Probability Theory and its Applications*. John Wiley & Sons Inc., New York.

Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*. Academic Press Inc., Toronto.

Galloway, M.M., 1975. Texture Analysis Using Gray Level Run Lengths. *Computer Graphics and Image Processing*, 4:172–179.

Gonzalez, R.C., and P. Wintz, 1987. *Digital Image Processing*. Addison-Wesley Publishing Company. New York. 503 p.

Haralick, R.M., 1979. Statistical and Structural Approach to Texture. *IEEE Proceedings*, 67(5):786–804.

Haralick R.M., K. Shanmugan, and I. Dinstein, 1973. Textural Features for Image Classification. *IEEE Transaction on Systems, Man and Cybernetics*, 6:610–621.

Holmes, Q.A., D.R. Nuesch, and R.A. Shuchman, 1984. Textural Analysis and Real-Time Classification of Sea-Ice Types Using Digital SAR Data. *IEEE Transaction on Geoscience and Remote Sensing*, GE-22(2):113–120.

Jha, R., and M.E. Jernigan, 1990. Classification and Segmentation of SAR Sea Ice Imagery Using Stochastic Image Models. *IGARSS'90*, 3:1881–1884.

Kwok, R., E. Rignot, B. Holt, and R. Onstott, 1992. Identification of Sea Ice Types in Spaceborne Synthetic Aperture Radar Data. *Journal of Geophysical Research*. 97(C2):2391–2402.

Livingstone, Charles E., R.G. Onstott, L.D. Arsenault, A. Laurence Gray, and Keshava P. Singh, 1987. Microwave Sea-Ice Signatures Near the Onset of Melt. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(2):174–187.

Matsuyama, T., S.-I. Miura, and M. Nagao, 1983. Structural Analysis of Natural Textures by Fourier Transformation. *Computer Vision, Graphics, and Image Processing*, 24:347–362.

McKillop, R., J.F. Sykes, and E.D. Soulis, 1990. Arctic Sea Ice Classificaiton in SAR Imagery Using Texture. *Proceedings of the 13th Canadian Symposium on Remote Sensing*, Fredericton, New Brunswick.

Nichols, A.D., J.W. Wilhelm, T.W. Gaffield, R.D. Inkster, and S.K. Leung, 1986. A SAR for Real-Time Ice Reconnaissance. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(3):383–389.

Nystuen, J.A., and F.W. Garcia, 1992. Sea Ice Classification Using SAR Backscatter Statistics. *IEEE Transactions on Geoscience and Remote Sensing*. 30(3):502–509.

Pultz, T.J., and R.J. Brown, 1987. SAR Image Classification of Agricultural Targets Using First- and Second-Order Statistics. *Canadian Journal of Remote Sensing*, 13(12):85–91.

Rauste, Y., 1990. Use of Texture Features in Discrimination of Sea Ice Types in SAR Images. *IGARSS'90*, 3:2229–2232.

Reed, T., and H. Weschler, 1990. Segmentation of Textured Images

and Gestalt Organization Using Spatial/Spatial-Frequency Representations. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-12(1):2–11.

Roan, S.J., and J.K. Aggarwal, 1987. Multiple Resolution Imagery and Texture Analysis. *Pattern Recognition*, 20(1):17–31.

Rosenfield, G.H., and K. Fitzpatrick-Lins, 1986. A Coefficient of Agreement as a Measure of Thematic Classification Accuracy. *Photogrammetric Engineering & Remote Sensing*, 52(2):223–227.

Shanmugan, K.S., V. Narayanan, V.S. Frost, J.A. Stiles, and J.C. Holtzman, 1981. Textural Features for Radar Image Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, GE-19(3):153–156.

Shapiro, S., and A. Gross, 1981. *Statistical Modelling Techniques*. Marcel Dekker Inc., New York, pp. 241–246.

Shokr, M.E., 1990. On Sea-Ice Texture Characterization from SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, GE-28(4).

————, 1991. Evaluation of Second-Order Texture Parameters for Sea Ice Classification From Radar Images. *Journal of Geophysical Research*, 96(C6):10,625–10,640.

Shokr, M.E., and D.G. Barber, 1990. A Technique for Sea-Ice Texture Quantification from Synthetic Aperture Radar. *Proceedings of the 13th Canadian Symposium on Remote Sensing*, Fredericton, N.B., Canada.

Skriver, H., 1991. Sea Ice Type Concentration from MIZEX-87 SAR Data. *IGARSS'91*, 2:415–418.

Srivastava, M., and E. Carter, 1983. *An Introduction to Applied Multivariate Statistics*. North-Holland, New York.

Story, M., and R.G. Congalton, 1986. Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering & Remote Sensing*, 52(3):397–399.

Sun, C., and W.G. Wee, 1982. Neighbouring Gray Level Dependence

Matrix for Texture Classification. *Computer Vision, Graphics and Image Processing*, 23:341–352.

Thomas, R.W., 1981. *Information Statistics in Geography*. CATMOG 31, Concepts and Techniques in Geography, Geo Abstracts, University of East Anglia, Norwich, 42 p.

Ulaby, F.T., F. Kouyate, B. Brisco, and T.H. Lee Williams, 1986. Textural Information in SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 24(2):235–245.

Unser, M., and M. Eden, 1989. Multiresolution Feature Extraction and Selection for Texture Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11(7):717–728.

Wackerman, C., 1991. Optimal Linear Combinations of Statistics and Texture Measures for SAR Sea Ice Classification. *IGARSS'91*, 1:113–116.

Wells, D., and D. Haygen, 1991. *A Knowledge Base to Support SAR Image Interpreation Systems for Arctic Ice Conditions*. Unpublished report by Norland Science and Engineering Ltd for the Applications Division Canada Centre for Remote Sensing, Ottawa, Ontario, 55 p.

Weszka, J. S., C.R. Dyer, and A. Rosenfeld, 1976. A Comparative Study of Texture Measures for Terrain Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(4):269–285.

Wilson, R., and M. Spann, 1988. Finite Prolate Spheroidal Sequences and Their Applications II: Image Feature Description and Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-10(2):193–203.

Wright, G., 1988. *Feature Selection for Texture Discrimination*. Unpublished M.A.Sc. Thesis, Department of Systems Design Engineering, University of Waterloo, Ontario.