ARKTOS: An Intelligent System for SAR Sea Ice Image Classification

Leen-Kiat Soh, *Member, IEEE*, Costas Tsatsoulis, *Senior Member, IEEE*, Denise Gineris, and Cheryl Bertoia

Abstract—We present an intelligent system for satellite sea ice image analysis named Advanced Reasoning using Knowledge for Typing Of Sea ice (ARKTOS). ARKTOS performs fully automated analysis of synthetic aperture radar (SAR) sea ice images by mimicking the reasoning process of sea ice experts. ARKTOS automatically segments a SAR image of sea ice, generates descriptors for the segments of the image, and then uses expert system rules to classify these sea ice features. ARKTOS also utilizes multisource data fusion to improve classification and performs belief handling using Dempster–Shafer. As a software package, ARKTOS comprises components in image processing, rule-based classification, multisource data fusion, and graphical user interface-based knowledge engineering and modification. As a research project over the past ten years, ARKTOS has undergone phases such as knowledge acquisition, prototyping, refinement, evaluation, deployment, and operationalization at the U.S. National Ice Center. In this paper, we focus on the methodology, evaluations, and classification results of ARKTOS.

Index Terms—Data fusion, Dempster–Shafer belief theory, intelligent image analysis, rule-based system, sea ice classification.

I. INTRODUCTION

REMOTE sensing of the polar regions has important applications in meteorology and in global climate studies. For example, the thickness of sea ice influences the heat flux between the atmosphere and water surface; thus, the classification and temporal tracking of sea ice can be used as an indicator in global climate monitoring [1]. Increased concern regarding global climate change and the subsequent increase in the number of earth-orbiting satellites have resulted in a dramatic increase in the volume of satellite imagery available to scientists. Thus, automation in sea ice image classification is desired to assist sea ice experts in extracting geophysical information from the increasing volume of images. Such automation releases sea ice experts from the task of having to retrieve and disseminate different sources of data to classify images, allowing them to concentrate on more important decision making.

There are many advantages to having a knowledge-based approach to sea ice image analysis. First, it mimics the reasoning process of sea ice experts and, thus, allows easier evaluation and knowledge refinement by experts. This close interaction also enables the software engineers and researchers to communicate with the experts using explicit knowledge. Second, it is convenient for multisource data fusion. Derived information can be readily added to the system with minimal programming impact, and new rules can be plugged into the knowledge base easily. Third, it is modular. Different knowledge bases may be built for images of different regions, different seasons, and different applications. Fourth, because of its modularity, many of the research and development processes have been conducted in parallel or in overlapping phases. A knowledge base that is stable can be promoted to be operational while another knowledge base may still be undergoing refinement. This equips the system with the ability to evolve cost-effectively. Finally, with knowledge explicitly represented and available for evaluation, this approach introduces accountability and encourages knowledge transfer and exchange among experts. Expert analysts may use the knowledge bases to train young analysts; experts may exchange their knowledge bases, using a similar language and subject particular rules to discussions and improvements; and users may understand why certain images are classified the way they are and may know which rules are the reasons behind the classification and which experts wrote the rules. This accountability enhances the knowledge-engineering culture within the sea ice community and makes sea ice expertise better organized and better portable.

There has been previous work in knowledge-based systems for remote sensing, such as aerial image understanding [2], land change detection [3], segmentation [4], and vegetation classification [5]–[7]. However, most knowledge-based systems are pixel-based, while the approach taken by Advanced Reasoning using Knowledge for Typing Of Sea ice (ARKTOS) is feature-based. Human experts do not analyze the images at the pixel level; instead, they look at regions and features and reason about them.

ARKTOS is also the only system that classifies synthetic aperture radar (SAR) sea ice data automatically in an operational environment, i.e., in near real-time. Other SAR sea ice classifiers include the work by Fetterer *et al.* [8], who developed the Multi-Year Ice Mapping System (MIMS) at the University of Colorado. MIMS was designed to quickly map old ice in uncalibrated SAR images using a local dynamic thresholding. MIMS only identifies old ice in high latitudes. The RADARSAT geophysical processor system (RGPS) classifies sea ice into local age and thickness distributions using ice motion and an empirical relation between accumulated freezing-degree days and ice thickness [9]. The RGPS requires occasional human interven-

TABLE I
FILE FORMATS AND BRIEF DESCRIPTIONS

File Format	Description
PGM	Portable GrayMap (PGM) is a standard bitmap based format consisting of a 4 lines header, and data stored in the *unsigned char* type, providing a maximum of 256 gray scale levels or 8-bit data per pixel.
GRIB	Data files in GRIdded Binary (GRIB) format with .GRB extension, an international standard.
NSIPS	Naval Satellite Image Processing System, a system that generates a customized data file format.

tion to identify tie points used in calculating ice motion, works only during the winter, is not near real-time, and cannot be used in an operational environment.

In the following, we first give an overview of ARKTOS. In Section III, we describe in detail the methodology of ARKTOS, including the various components. In Section IV, we present the evaluation results. Finally, in Section V, we summarize ARKTOS and make recommendations for improving the performance of the system.

II. OVERVIEW OF ARKTOS

ARKTOS is a fully automated intelligent classifier of sea ice in SAR images, which also incorporates and fuses ancillary data sources to support its classification conclusions. Given an image, ARKTOS extracts objects or *features* and then computes a set of attributes for each feature. Next, ARKTOS feeds the features with the corresponding attributes into a rule-based system. The rule-based system is supported by a Dempster–Shafer belief system [10]. Each rule has an antecedent, a consequent, and a weight. The antecedent consists of attribute–value pairs describing a feature in the image. The consequent is the resultant assertion that the feature belongs to a certain ice class (in the current version of ARKTOS the classes are open water (OW), new ice (NI), first-year (FY) ice, and multiyear (MY) ice). The weight is the confidence in the assertion. A feature may trigger the firing of multiple rules, asserting complimentary or conflicting ice classifications. The Dempster–Shafer belief system collects these weights as masses of evidence and combines them to compute the belief and plausibility of a feature belonging to a particular class.

ARKTOS performs multisource data fusion [11], [12] by integrating data of different formats and sources to help classify the features. Since these data are of different resolutions and domains, ARKTOS uses georeferenced conversions and attribute measurements to bring them to a common, usable form.

ARKTOS is also a knowledge-engineering tool [13], [14]. It consists of a suite of graphical user interfaces (GUIs) that allows users to refine the system and review the performance of the software, verifying the classification rules. Initially, we conducted knowledge acquisition from sea ice experts and then built a prototype quickly. This stage involved interviewing sea ice experts, transcribing the sessions, identifying descriptors and rules, designing and implementing the knowledge, and delivering a modestly accurate classification prototype quickly. A refinement stage then involved evaluating the prototype, refining the knowledge base, modifying the design, and reevaluating the improved system. Finally, ARKTOS transitioned to operations at the National Ice Center [15].

The ARKTOS software package operates on Windows NT and UNIX platforms, using Naval Satellite Image Processing System (NSIPS)-processed (by the Naval Research Laboratory) [16] SAR images, and fusing other data such as climatology, and Special Senor Microwave/Imager (SSM/I) ice concentration maps. ARKTOS was written in the C programming language except for the GUIs, which were written in Java.

III. METHODOLOGY

There are four main components in ARKTOS: image processing, rule-based classification, multisource data fusion, and a suite of JAVA-based GUIs that form the knowledge-engineering and evaluation component.[1]

A. Image Processing

We have studied, designed, and implemented four stages of the image processing sequence in the course of researching and designing ARKTOS: preprocessing, segmentation, attribute measurement, and fact generation. As a feature-based classification system, the attribute set of each feature in ARKTOS is matched against a set of rules. These rules are used to assert or refute evidence that a feature belongs to a particular ice class. Therefore, it is important for us to consider techniques at both the feature and global level. At the global level, we have the preprocessing and segmentation techniques applied to the entire image. At the feature level, we have the attribute measurements and symbolic description stages applied to each feature found.

1) Preprocessing: The SAR images that ARKTOS handles come in various formats, processed by four different satellite reception stations: the Alaska SAR Facility (ASF), Tromsø, Norway, Gatineau, Canada, and West Freugh, Scotland. In addition, there are images that have been preprocessed by the National Ice Center into NSIPS format. ARKTOS converts these images into PGM format with 256 gray levels (see http://netpbm.sourceforge.net/doc/pgm.html).

For a non-NSIPS-processed image, ARKTOS performs a 5×5 Gaussian sampling to reduce both the size of the image and noise. The Gaussian sampling is a weighted intensity average within a window using weights of a two-dimensional Gaussian curve.

The current version of ARKTOS uses several external information sources: SSM/I ice concentration GRIB files, landmasks, and two sets of ice climatology data, both extracted from the records of the NIC: one representing the probability of a region containing ice in a 19-year span and the other representing

---

[1]In this paper, we do not discuss the knowledge-engineering process, which is described in detail in [14].

TABLE II
FEATURE ATTRIBUTE MEASUREMENTS COMPUTED BY ARKTOS AND THEIR DESCRIPTIONS

Measurement	Description
Area	Measures the number of pixels within a feature.
Average_intensity	Measures the average intensity of all pixels within a feature.
Standard_deviation	Measures the standard deviation of the intensity values of all pixels within a feature.
Contrast	Measures the contrast of all pixels within a feature. $$contrast = \frac{standard\_deviation}{average\_intensity}$$ Suppose we are given two images with the same *standard_deviation*. If one of the images is bright, with a high *average_intensity*, and the other is dark, with a low *average_intensity*, then the darker image has a higher *contrast* value. In a sense, we expect a brighter image to have a larger standard deviation in order to have a high contrast.
Perimeter	Measures the number of boundary pixels of a feature.
Outer_perimeter	Measures the number of boundary pixels when traversing the outer boundary of a feature. This number may be larger or smaller than *perimeter*. When a feature contains holes, *perimeter* is greater than *outer_perimeter*. When a feature has one-pixel-thick structures, *outer_perimeter* is greater than *perimeter*.
Perimeter_porosity	This measurement allows us to determine if a feature has many holes and 1-pixel-thick structures. $$perimeter\_porosity = \frac{\max(outer\_perimeter, perimeter)}{\min(outer\_perimeter, perimeter)}$$
Centroid	Stores the pixel coordinates $\langle \mu_x, \mu_y \rangle$ of the center of mass of a feature.
Mottledness	The objective of this measurement is to detect the intensity differences in a feature. If we detect high intensity differences, then we say the feature is mottled. Also, we tolerate less difference for bright features than we do for dark features to capture the observation that a slight change at the bright end of the intensity spectrum is more significant than one at the dark end. (1) Obtain the maximum difference between two horizontal, neighboring pixels of the feature, $\max_{horizontal}$. (2) Obtain the maximum difference between two vertical, neighboring pixels of the feature, $\max_{vertical}$. (3) Compute $$mottledness = (\max_{vertical} + \max_{horizontal}) \bullet \frac{average\_intensity}{255}.$$ The value 255 is used as a normalization factor.
Average_roughness	This is a new approach to measuring the texture of a feature by obtaining a mean of variances of local regions of a feature. (1) Divide the feature into overlapping 5x5 areas. (2) Compute the variance of each area. (3) Sum all the variances. (4) Divide the sum with the feature's area.
New_roughness	This is yet another new approach to measure the texture of a feature by combining a feature's global and local variances . This measure is computed as $$new\_roughness = \frac{standard\_deviation^2}{average\_roughness}$$ Suppose a feature has a low *average_roughness* but a high *standard_deviation*. That means locally, the feature is uniform; but globally, the feature has a larger intensity range and wider intensity distribution. This implies two possible conditions: (1) the feature has a significant shift in intensity values from one part of the feature to another part, or (2) the feature is patchy (for example, a checkerboard pattern).

the median concentration of ice in that region during the same span [17]. ARKTOS converts all these datasets into PGM format for faster access at run-time.

Table I shows a brief summary of the different file formats.

2) Segmentation: We studied three main segmentation algorithms: 1) dynamic local thresholding, 2) unsupervised clustering, and 3) watershed merging. Dynamic local thresholding is not suitable because it segments an image into n classes based on their global appearance, instead of feature-level homogeneity [18]. Unsupervised clustering is an aggressive pixel aggregation technique that is not suitable for feature extraction, as it merges too many features into one single region [19]. Watershed merging was deemed the best segmentation technique for our purposes and was implemented in ARKTOS [20], [21]. In geography, watersheds are regions of terrain that drain toward the same point. This situation can be analogously applied to SAR sea ice images by treating intensity as height. First, the algorithm identifies the local intensity minima that define the bottoms of watersheds. A minimum is defined as a pixel with all its eight neighbors having greater intensities than the pixel. Then, the algorithm computes the image gradient and partitions the input image into watersheds by marking the locations of intensity minima with unique region identifiers in an output image. For each of the remaining pixels, the gradient information is used to follow the image down to some intensity minimum. The corresponding pixel location in the output image is assigned the identifier of this minimum. The watershed merging algorithm subsequently merges each pair of neighboring watersheds based on their average intensities, sizes, and gradients.

TABLE II
(Continued.) FEATURE ATTRIBUTE MEASUREMENTS COMPUTED BY ARKTOS AND THEIR DESCRIPTIONS

Orientation	This is used to gauge how a feature aligns in the image and is useful for computing the feature's bounding rectangle, as will be discussed later. $$orientation = \frac{1}{2}\tan^{-1}\left[\frac{2\bar{\mu}_{1,1}}{\bar{\mu}_{2,0} - \bar{\mu}_{0,2}}\right]$$ where $$\bar{\mu}_{0,2} = \sum_{all\,y}(y - \mu_y)^2, \quad \bar{\mu}_{2,0} = \sum_{all\,x}(x - \mu_x)^2, \text{ and } \bar{\mu}_{1,1} = \sum_{all\,x}\sum_{all\,y}(x - \mu_x)(y - \mu_y).$$
max_length and max_width	The approach is to compute the bounding rectangle of a feature, and use the two principal axes to find the maximum length and width of the feature. (1) For each pixel $\langle x, y\rangle$, compute its $$\alpha = x\cos(orientation) + y\sin(orientation)$$ $$\beta = -x\sin(orientation) + y\cos(orientation)$$ (2) Identify the maximum α as α_{max}; identify the maximum β as β_{max}; identify the minimum α as α_{min}; identify the minimum β as β_{min}. (3) Compute the distances of the principal axes: $$d_\alpha = \alpha_{max} - \alpha_{min}, \text{ and } d_\beta = \beta_{max} - \beta_{min}.$$ (4) Compute $$max\_length = \max(d_\alpha, d_\beta), \text{ and } max\_width = \max(d_\alpha, d_\beta).$$
Area_porosity	This measurement allows us to determine if a feature is branchy. $$area\_porosity = \frac{max\_width \bullet max\_length}{area}$$
Elongation	Measures how elongated a feature is. $$elongation = \frac{max\_length}{max\_width}$$
Roundness	Measures how circular a feature is. First, for each *boundary* pixel, p, we compute its distance from the centroid $\langle\mu_x, \mu_y\rangle$: $$D(p, \langle\mu_x, \mu_y\rangle) = \sqrt{(x_p - \mu_x)^2 + (y_p - \mu_y)^2}.$$ Then we compute the standard deviation of all such distances and equal roundness to the standard deviation: $$roundness = std\_dev[D(p, \langle\mu_x, \mu_y\rangle)].$$
Thinness	Measures how thin a feature is. We have designed a quick and easy way to approximate elongation without having to consider the orientation of a feature. The underlying concept of this approach is to estimate the average thickness of a structure. (1) Scan the image horizontally, collect the length at each row of the contiguous segment of the feature, and average all lengths. Call the average $\bar{l}_{horizontal}$. (2) Scan the image vertically, collect the length at each column of the contiguous segment of the feature, and average all lengths. Call the average $\bar{l}_{vertical}$. (3) Compute *thinness* as: $$thinness = \min(\bar{l}_{horizontal}, \bar{l}_{vertical}).$$
Irregularity	Measures how irregular a feature is: $$irregularity = area\_porosity \bullet perimeter\_porosity.$$
Eccentricity	Measures how eccentric a feature is. A feature is eccentric if it has boundary pixels that are very close to its centroid and boundary pixels that are very far from its centroid. A circle is not eccentric because all boundary pixels are equidistant from the centroid. An

Specifically, the watermerge algorithm works as follows. After the initial watershed-based segmentation, the image is divided into a set of regions. Each region is attributed with an average intensity and a size. The boundary gradient between each pair of regions is also computed. In addition, there are three basic thresholds: a boundary gradient threshold, $Thresh_g$, which is set at 6.0, an average intensity threshold $Thresh_i$, which is set at 12, and a size threshold $Thresh_a$, which is set at 10. For each image, we adapt the boundary gradient threshold to $Thresh_g' = Thresh_g \cdot (range/255)$, where range is the dynamic range of the image. Similarly, we adapt the average intensity threshold to $Thresh_i' = Thresh_i \cdot (range/255)$. Next, the watermerge algorithm performs two layers of merging.

The first layer carries out ten iterations of merging based on the boundary gradient between each pair of regions. During iteration i, we merge a pair of regions if the boundary gradient between the two is less than $(i \cdot Thresh_g')/10$, and the size of the smaller region is $(i \cdot Thresh_a)/10$. When the algorithm arrives at the last iteration, it merges a pair of regions if the boundary gradient between the two is less than $Thresh_g'$ and the size of the smaller region is $Thresh_a'$. The strategy here is to merge smaller regions with smaller gradient differences first, before moving to the next iteration. During the next iteration, newly merged regions will be considered for further merging. This strategy has been shown to perform well even with range effects in images [20].

TABLE II

(Continued.) FEATURE ATTRIBUTE MEASUREMENTS COMPUTED BY ARKTOS AND THEIR DESCRIPTIONS

	N-pointed star is eccentric, however, because the boundary pixels at its points are farther away from the centroid than the pixels at the valley between points. This attribute is used to complement *irregularity*. First, for each *boundary* pixel, p, we compute its distance from the centroid $\langle \mu_x, \mu_y \rangle$: $$D\left(p, \langle \mu_x, \mu_y \rangle\right) = \sqrt{\left(x_p - \mu_x\right)^2 + \left(y_p - \mu_y\right)^2} .$$ Then, we identify the maximum distance, d_{max}, and the minimum distance, d_{min}. Finally, $eccentricity = d_{max}/d_{min}$.
Jaggedness	Measures how jagged the boundary of a feature is. As we traverse the boundary, we take the absolute difference between the previous direction that brought us to the current pixel and the current direction that brings us to the next pixel. In a 3×3 8-connected neighborhood, the maximum change in directions is 4; the minimum is 0. As we move along the boundary, we add the magnitude of each directional difference to a counter. Jaggedness is defined to be the average of all such direction changes, i.e., the value of the counter divided by outer_perimeter.

Neighbor_intensity	Measures the average intensity of the neighbors weighted by the number of shared boundary pixels.
	(1) Follow the boundary of a feature
	    (2) For each boundary pixel encountered
	        (3) Look at is right-hand (outer) neighboring pixel
	        (4) Identify the feature number of that pixel
	        (5) Retrieve the average_intensity associated with the feature number
	        (6) Add the value to a counter
	(7) Divide the counter with outer_perimeter.
	So, suppose a feature has two neighbors. One is extremely dark, the other extremely bright. However, the dark neighbor shares boundary pixels with 90% of the feature's boundary while the bright neighbor shares only 10%. Then, as a result, neighbor_intensity will record a low value. If we treated the two neighbors equally (50% - 50%), then neighbor_intensity would record a medium value, which would be undesirable.

Neighbor_mottledness	This is very similar to neighbor_intensity, but looks at the mottledness instead of intensity values of the neighboring features.
Curved_boundary	Boundary-based measurements are expensive, as we have to traverse along the boundary of the feature. Thus, we have created three approximation measurements. Curved_boundary is one of them. This detects whether a feature has a curved (or rounded) boundary. We first obtain five successive, equally long segments along the boundary and compute the tangent of each segment. If (1) all angles are monotonically increasing, and (2) the angle change between each pair of successive segments is smaller than a threshold, $T_{curved,0}$, and (3) the total angle change is greater than a threshold, $T_{curved,1}$, then we say that the feature has a curved boundary, i.e., curved_boundary = 1. If the current group of five segments does not amount to a curved boundary; we discard the first segment, include the next segment along the boundary into the group, and repeat the above process. If after traversing the boundary *twice* we find no curved boundary, then we set curved_boundary = 0. The first condition is to discard boundary segments that move inward and outward, with respect to the centroid of the feature. The second condition is to ensure smooth angle transition since it does not allow sharp jump in the boundary. The final condition is to ensure a significant curvature. The slope of each segment is computed by taking the slope between the two ends of the segment: $$slope_i = \frac{y_{end} - y_{start}}{x_{end} - x_{start}}$$

The second layer also carries out ten iterations of merging, but based on the average intensity difference between each pair of regions. During iteration i, we merge a pair of regions if average intensity difference between the two is less than $(i \cdot \text{Thresh}'_i)/10$ and the size of the smaller region is $(i \cdot \text{Thresh}_a)/10$.

Note that in the first layer, the algorithm focuses on the strength of the boundary. If the boundary is not strong enough, then the two regions are merged. In the second layer, even if the boundary is strong enough, it is possible that from a global viewpoint, the average intensity of the two regions is very similar. If they have similar average intensity values, they are merged.

3) Feature Attribute Measurements: The segmentation stage identifies intensity-based homogeneous regions in the image as features. For each feature, we compute a set of attribute measurements that helps us generate the facts that are needed by the classification rules. The set of attributes used to define a feature was based on conversations with and knowledge acquisition from expert analysts of SAR sea ice imagery. They attempt to capture what the experts defined as the visual cues they use when classifying sea ice. Some of these attributes are common sense (e.g., the area or average intensity of a feature), while others are domain-specific and have names assigned to them by the experts (e.g., "mottledness" as a measure of texture variation of a feature). Since ARKTOS needs to classify images quickly, some traditional image analysis measurements, such as grayscale cooccurrence matrices, were not used, as they are computationally expensive. The complete

TABLE II
(Continued.) FEATURE ATTRIBUTE MEASUREMENTS COMPUTED BY ARKTOS AND THEIR DESCRIPTIONS

	where $\langle x_{start}, y_{start} \rangle$ and $\langle x_{end}, y_{end} \rangle$ are the start- and end-point of the segment i.
Linear_boundary	This is another boundary-related measurement and detects whether a feature has a linear boundary. We divide the boundary into several segments with length proportional to the size of the boundary, or outer_perimeter. For each segment, we count the number of image border pixels. If this number is greater than 30% of the length of the segment, we do not perform linear regression on this segment and it is discarded and we proceed to the next segment. If, however, the segment contains image border pixels less than 30% of its length, then we perform a linear regression on the segment. First we compute the centroid of the segment (similar to the centroid of the feature), $\langle s_x, s_y \rangle$. Then we calculate the moments: $$s_{yy} = \sum_{all\,y}(y - s_y)^2 \;,\; s_{xx} = \sum_{all\,x}(x - s_x)^2 \;,\; \text{and}\; \mu_{xy} = \sum_{all\,x}\sum_{all\,y}(x - s_x)(y - s_y)\;.$$ We compute the coefficient of determination, η, as $$\eta = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}\;.$$ If η is greater than a threshold, T_{linear}, then we have found a linear segments, i.e., *linear_boundary* = 1. Otherwise, we proceed until we run out of segments. If we do not find such a segment, then the feature has *linear_boundary* = 0.
Angular_boundary	This is the third boundary-related measurement and it detects whether a feature has an angular boundary. The underlying design is very similar to that for linear_boundary. Instead of looking for one linear segment, we are looking for two consecutive linear segments. After we locate the first linear segment, we continue to extend that segment as long as it stays linear by checking at every move along that segment the coefficient η. At the end of that first linear segment, η will be small (no longer linear). Then we immediately look at the next segment. If this segment is found to be linear, then we have found two successive linear segments, and angular_boundary = 1. If the segment is not linear, then we discard the first linear segments and continue the search. If we run out of segments and no two successive linear segments are found, angular_boundary = 0.

set of the attributes we compute and their descriptions are listed in Table II.

The first set of attributes we measure are intrinsic and include area, average intensity, standard deviation of the intensity values in the feature, and contrast (the ratio of the standard deviation over the average intensity).

To measure boundary-related attributes, we first build the chain code of a feature. The chain code has a starting point and a set of directions leading from that starting point and ending at the starting point. It is an efficient way of storing boundary information and traversing a boundary. We then compute the length of the boundary (or perimeter) and the length of only the outer boundary where the perimeter length of internal holes of an object is not included. We use these to compute perimeter porosity, which is simply the ratio of the greater perimeter over the shorter one. This attribute is used to specifically describe the degree of irregularity of sea ice features. A sea ice feature may also have different boundary types, such as curved, linear, or angular; these boundary types are sometimes good proxies of a feature's age.

The third type of attributes is texture-based. These turned out to be very difficult to design and implement, as human experts are able to detect complex textures that computer algorithms fail to capture accurately. Currently, we have three different types of textural attributes: mottledness, average roughness, and new roughness. The objective of mottledness is to detect the intensity differences within a feature: "mottled" features display high-intensity differences. Based on interviews with experts, "mottledness" is designed to tolerate less difference for bright features than for dark features in order to capture the experts' observation

that a slight change at the bright end of the intensity spectrum is more significant than one at the dark end. The average roughness is a more traditional, but far more time-consuming, texture measure, in that it uses overlapping 5×5 windows to compute variances and then averages them over the feature. New roughness was designed specifically by the experts for sea ice features, and integrates average roughness with the standard deviation of the intensity of the feature. The goal is to capture large-scale textual variations, where locally (in a 5×5 window) there is homogeneity but globally there is texture variability.

To obtain geometric attributes, we compute principal axes, the centroid, and the bounding rectangle of each feature. Given these, we are able to determine the orientation of a feature, its maximum length and width, area porosity, elongation, roundness, thinness, irregularity, eccentricity, and jaggedness. Most of these attributes are specifically designed for sea ice features. For example, the area porosity of a feature is the ratio of the bounding rectangle over the actual area of the feature and approximates the "branchiness" of the feature. A feature with branches has a high area porosity, and this corresponds to ice leads, a rather unique feature observed in sea ice images. Also, a feature is eccentric if it has boundary pixels that are close to its centroid and boundary pixels that are far from its centroid. A circle is not eccentric, because all boundary pixels are equidistant from the centroid. An *N-pointed* star is eccentric, however, because the boundary pixels at its *points* are farther away from the centroid than the pixels at the valley between points. Irregularity is an innovative complex attribute: a feature that has a high perimeter porosity (a lot of holes) and a high area porosity (a lot of branches) is highly irregular.

TABLE III
FACTS EXTRACTED FROM THE FEATURE MEASUREMENTS FROM THE SAR IMAGE

Fact	Possible Discrete Values	Feature Measurements that are Quantized	Thresholds Used
return	black, dark, gray, bright	Based on the measurement average_intensity. If $average\_intensity < T_{return,0}$, then return=black. If $T_{return,0} \leq average\_intensity < T_{return,1}$, then return=dark. If $T_{return,1} \leq average\_intensity < T_{return,2}$, then return=grey. If $T_{return,2} \leq average\_intensity$, then return=bright.	$T_{return,0} = 50$, $T_{return,1} = 75$, $T_{return,2} = 100$
size	small, medium, large	Based on the measurement area. There are three thresholds related to this attribute; however, only two are used to d iscretized and another one is used to qualify blobs (discussed later). If $area < T_{size,0}$, then size=small. If $T_{size,0} \leq area < T_{size,1}$, then size=medium. If $T_{area,1} \leq area$, then size=large.	$T_{size,0} = 200$, $T_{size,0} = 1600$, $T_{size,0} = 25000$
mottled or smooth	mottled=true, smooth=true	Based on the measurement mottledness. If $mottledness > T_{mottled}$ then mottled=true; else smooth=true. Instead of mottled=false, we use smooth=true because the latter is more intuitive.	Observed values between 5.0 and 60.0. $T_{mottled} = 31.0$
blob	true, false	Based on the measurements area, irregularity, and eccentricity. We use three thresholds, related to area, irregularity, and eccentricity: If $area > T_{size,2}$ and $(irregularity > T_{irregular}$ or $eccentricity > T_{eccentric})$ then blob=true; else blob=false.	$T_{size,2}$, $T_{irregular}$, $T_{eccentric}$
round	true, false	Based on the measurement roundness. If $roundness < T_{round}$ then round=true; else round=false.	A perfect circle has a measurement of 0. $T_{round} = 1.05$
elongated	true, false	Based on the measurement elongation. If $elongation > T_{elongated}$ then elongated=true; else elongated=false.	Observes values between 1 and 4. $T_{elongated} = 1.3$
irregular	true, false	This denotes whether a feature has an irregular shape. In addition to irregularity, we also use eccentricity to help define this attribute. If $(irregularity > T_{irregular}$ or $eccentricity > T_{eccentric})$ then irregular=true; else irregular=false.	Observed values between 1.0 and 25.0. $T_{irregular} = 3.10$ Observed values between 2.0 and 20.0. $T_{eccentric} = 4.50$

Note that our watershed-based segmentation technique may sometimes merge regions overaggressively, resulting in a feature of irregular shape. By describing these features with our irregularity measurement, ARKTOS singles them out and classifies them using specific rules. Thus, ARKTOS is able to compensate for segmentation overmerging.

4) Fact Generation: The feature attribute measurements (except those for curved, angular, and linear boundaries) are continuous, real-valued numbers that cannot be used in their raw form for classification. Human analysts express their knowledge in a discrete way. For example, a feature is of either "small," "medium," or "large" size, or it is either elongated or not. Consequently, the feature attribute values had to be quantized and translated into a set of discrete values. In addition, other, higher level facts (such as a feature being a "lead") are generated. Thresholds were defined with the help of experts that divide continuous measurements into discrete fact values. Table III shows all facts with their discrete values, the attribute values that lead to them, and the expert-defined threshold(s). Some of the facts are relational, obtained by looking at the shared boundaries of the features: A encloses B; A is neighbor of B; A contains cracks; A is adjacent to land, etc. Other attributes are obtained by comparing attributes of features with shared boundaries: A is darker than its neighbors; A is more mottled than its neighbors, etc. These are important indicators in sea ice image analysis. For example, if a feature is found to be brighter than its neighbors, then it is more likely to be a piece of multiyear ice.

Most facts are self-explanatory except, possibly, for one: "blob." In some images, there appear features that are vast in size and basically featureless (often large areas of multiyear ice or open water); these are defined as "blobs." If a feature is found to be a blob, then all its geometric and boundary-related attributes (such as roundness, elongation, irregularity, etc.) are neither computed nor used in classifying the feature. This

TABLE III
(*Continued*). FACTS EXTRACTED FROM THE FEATURE MEASUREMENTS FROM THE SAR IMAGE

thin	true, false	Based on the measurement thinness.
		
		`If` *thinness* $< T_{thin}$ `then thin=true; else`
		`thin=false.`
			Observed values between 5.0 and 30.0.
			$T_{thin} = 11.0$

jagged	true, false	This denotes whether the boundary of a feature is jagged. It is based on the measurement jaggedness.
		
		`If` *jaggedness* $> T_{jagged}$ `then jagged=true; else`
		`jagged=false.`
			Ranges between 0 and 4.
			$T_{jagged} = 0.74$

lead	true, false	This indicates whether a feature is a lead. To be a lead, a feature must be elongated and irregular. Thus, this attribute is based on the measurements elongation and irregularity. Two thresholds discussed previously will be used here. `If` *elongation* $> T_{elongated}$ `and` *irregularity* $> T_{irregular}$ `then lead=true; else lead=false.`	$T_{irregular} = 3.10$, $T_{elongated} = 1.3$

enclose	true, false, darker, brighter	This denotes whether a feature encloses at least one other feature. It also denotes when a feature encloses something darker in average intensity. This is based on all neighbors of the feature of interest.
		`If the feature encloses nothing, then en-`
		`        close=false.`
		`If the feature encloses something, then`
		`        enclose=true.`
		
		`If the feature encloses something,` f_e `, and`
		$average\_intensity(feature) > \lambda \bullet average\_intensity(f_e)$
		`then enclose=darker.`
		
		`If the feature encloses something,` f_e `, and`
		$\lambda \bullet average\_intensity(feature) < average\_intensity(f_e)$
		`then enclose=brighter.`
		The factor λ is empirically set at 1.2.
			·

contain_cracks	true, false	This denotes whether a feature encloses a crack. A crack is defined as a feature that is elongated and thin. So, the thresholds discussed above for attribute elongated and thin will be utilized again.
		
		`If the feature encloses something,` f_e `, and`
		$elongation(f_e) > T_{elongated}$ `and`
		$thinness(f_e) < T_{thin}$ `then contain_cracks=true;`
		`        else contain_cracks=false.`
			$T_{elongated} = 1.3$,
			$T_{thin} = 11.0$

smoother	true, false	This denotes whether a feature is smoother than its surroundings. It is based on the measurement mottledness.
		`If mottledness of the feature is less than`
		`the average mottledness of all its`
		`neighbor then smoother=true; else`
		`smoother=false.`	

brighter	true, false	Based on the measurement average_intensity.
		`If average_intensity of the feature is`
		`greater than` λ `times the average aver-`
		`age_intensity of all its neighbors then`
		`brighter=true; else brighter=false.`	

important distinction is made, since rules are designed to describe individual ice floes rather than groups of ice types.

Figs. 1 and 2 illustrate facts generated by ARKTOS in relationship to their spatial (intensity and mottledness, in this case) and shape-related (roundness, elongation, and irregularity) attribute values. Fig. 1(a)–(c) gives examples of features of varying intensity, while Fig. 1(d)–(f) provides examples of features of varying mottledness. It should be noted that Fig. 1(e) represents a feature with a mottledness value just below the mottledness threshold of 15.00, which separates features that ARKTOS considers smooth from features that ARKTOS considers mottled. Fig. 2(a)–(c) provides examples of features of varying ARKTOS roundness, with Fig. 2(a) being the most round and Fig. 2(c) the least round. For this attribute, the smaller the value, the rounder the feature. Features with roundness values below the threshold of 1.05 are considered to be round

by ARKTOS, while those with values greater than 1.05 are not. Fig. 2(d)–(f) illustrate how the attributes of elongation and irregularity are considered together to characterize a feature as a lead. For a feature to be characterized as a lead, both its elongation value and its irregularity value must be above preset thresholds. The features in Figs. 2(d) and 2(e) have values that are all above the thresholds and are thus considered by ARKTOS to be leads. The feature in Fig. 2(f) has an elongation value higher than the threshold but not irregularity, so it is not considered to be a lead by ARKTOS.

B. Multisource Data Fusion

In addition to the raw image data, ARKTOS incorporates ancillary data into its reasoning process. This multisource data fusion component is necessary, as various data sources are often

TABLE III
(Continued). FACTS EXTRACTED FROM THE FEATURE MEASUREMENTS FROM THE SAR IMAGE

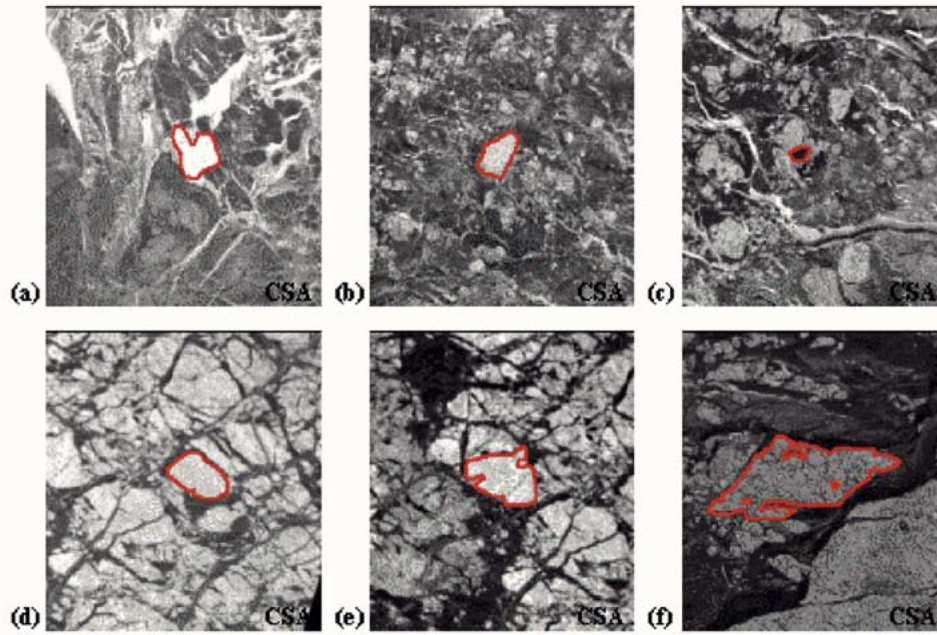|		

smoother2	true, false	Based on the measurement mottledness and neighbor_mottledness.
		`If mottledness of the feature is less than`
		`neighbor mottledness of the feature then`
		`smoother=true; else smoother = false.`
		The difference between smoother and smoother2 is that smoother treats each neighbor equally while smoother2 weighs each neighbor by the shared boundary between each neighbor and the feature.

brighter2	true, false	Based on the measurement average_intensity and neighbor_intensity.
		`If  average_intensity of the feature is`
		`greater than` λ `times neighbor_ intensity`
		`of  the feature then brighter= true; else`
		`brighter=false.`
		The difference between brighter and brighter2 is that brighter treats each neighbor equally while brighter2 weighs each neighbor by the shared boundary between each neighbor and the feature. The factor λ is the same one as in attribute enclose.

Fig. 1. Examples of spatial attribute values. Features with intensity values of (a) 162, (b) 84, (c) 36, where (a) is considered by ARKTOS to be bright, (b) gray, and (c) dark. Features with mottledness values of (d) 10.67, (e) 14.68, (f) 34.39, where (d) and (e) are considered by ARKTOS to be smooth and (f) mottled. Images are from November 7, 1999, March 29, 1998, and December 8, 2000. Images CSA 1998–2000.

used by human experts to help them classify sea ice images. Our fusion framework is attribute- and knowledge-based. The information derived from various data sources is converted into attributes and facts linked to each feature. Then, the knowledge on how to use these pieces of information is encoded in the rules of the rule-based classification module. The fusion is performed at two levels. At the attribute level, for each feature we find a combined list of attribute measurements computed from the imagery data and all other sources of data. For each data source, ARKTOS has to perform appropriate georeferencing and conversions. At the knowledge level, we define how different sources of data should work with each other through weighted rules and a Dempster–Shafer belief system (Section III-C). Currently, there are four different sources: an-

cillary data, SSM/I concentration maps, landmasks, and historical ice climatology data.

Ancillary data comes with the raw SAR image as the header, trailer, or leader information and specifies the date and the coordinates of where the image was taken. These data are integrated into the database through several facts:[2] summer $=$ true, winter $=$ true, west_of $=$ xxxx, east_of $=$ xxxx, and latitude $<=$ xxxx, where xxxx is a coordinate value in degrees. Each SAR image processing facility generally has a different ancillary data format to accompany its images. Ancillary data are also used to establish intersource information coordination

---

[2]Table IV shows all facts generated by ancillary data and that are used in the reasoning of ARKTOS.
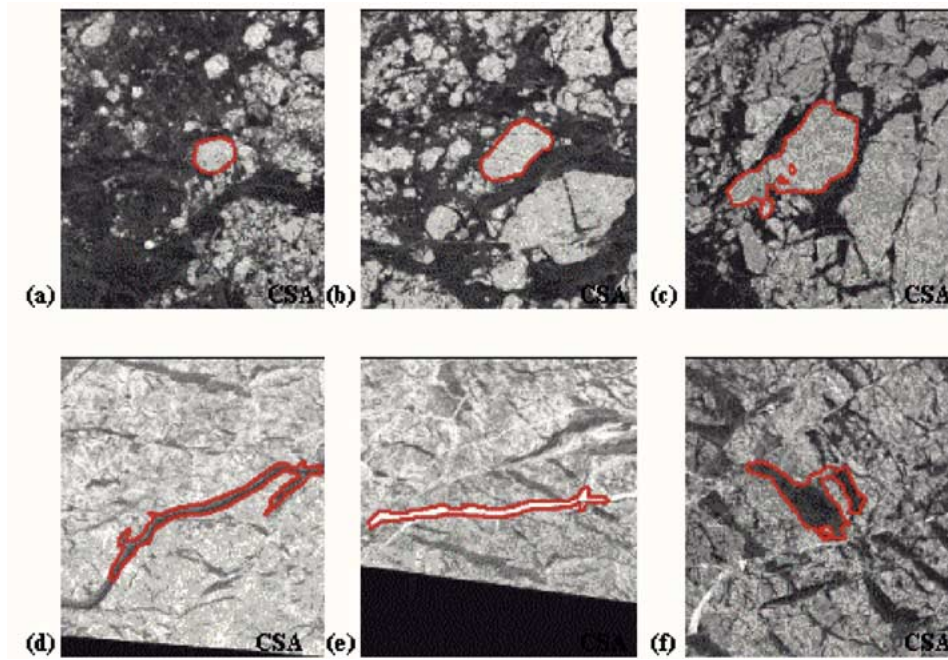
Fig. 2. Examples of shape attribute values. Features with roundness values of (a) 0.37, (b) 1.06, (c) 1.78, where (a) is considered by ARKTOS to be round, and (b) and (c) are not. Features with elongation and irregularity values of (d) 4.36 and 4.09, (e) 10.12 and 3.62, (f) 1.53 and 2.68, where (d) and (e) are considered by ARKTOS to be leads, and (f) is not. Images are from March 5, 2000, March 9, 2000, and March 12, 2000. Images CSA 2000.

TABLE IV
FACTS EXTRACTED FROM THE ANCILLARY DATA SOURCES

Fact	Possible Discrete Values	Description
adj_to_land	true, false	Depends on whether a feature touches land or not
ssmicon	low, med, high	These thresholds help qualify the attribute ssmicon, a value derived from the SSM/I sea ice concentration maps. Currently, these thresholds are set at 15.0 and 50.0, respectively. $T_{ssmicon,0} = 15.0$, $T_{ssmicon,1} = 50.0$
numyr	none, some, most, not_available	These thresholds help qualify the attribute numyr, a value derived from the climatology data ice coverage map. $T_{numyr,0} = 5$, $T_{numyr,1} = 15$
medct	low, med, high, not_available	These thresholds help qualify the attribute medct, a value derived from the climatology data ice coverage map. $T_{medct,0} = 20$, $T_{medct,1} = 75$
month	jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec	
season condition	summer, winter, freeze_up, , melt_out	Date thresholds provided by the experts
latitude	<=72.00, >=72.00, <=73.00, >=73.00, <=74.00, >=74.00, <=75.00, >=75.00	

and to select the corresponding ice chart and SSM/I concentration map, for example.

We also keep track of the geographic location (latitude and longitude coordinates) of the four corners of the image, key information for our fusion. Since we have attribute-level fusion, we need to tag each feature with its latitude–longitude location so that ARKTOS can locate the corresponding data point when moving from one data source to another. ARKTOS can analyze images using SSM/I sea ice concentration maps. These gridded

TABLE V
COMBINATION OF TWO MASS VALUES USING
DEMPSTER'S RULE OF COMBINATION

		{old_ice}	0.7	Θ	0.3
{open_water}	0.2	{}	0.14	{open_water}	0.06
Θ	0.8	{old_ice}	0.56	Θ	0.24

ice concentrations are generated with the calibration/validation (CAL/VAL) algorithm [22] in GRIB format by the U.S. Navy's

Rules
rule=4;If lead and return is dark then open_water;winter true,lead true,return dark;open_water;0.4
rule=6;If it is June and feature is very dark (black) then open_water;winter true,jun true,return black;open_water;0.7
rule=28;If the boundary is linear then first_year_ice;winter true,linear_boundary true;first_year_ice;0.5
rule=41;If the ice touches the land then it is land fast_ice;winter true,adj_coast true;fast_ice;0.6
rule=81;if it is small, dark, and darker than it could be FY;winter true,size small,return dark,brighter false;first_year_ice;0.8
rule=87;if not blob but irregular, thin, dark, darker than, and smoother than, then FY;winter true,blob false,irregular true,return dark,thin true,brighter false,smoother true;first_year_ice;0.7
rule=95;if the SSM/I concentration is low, then the feature is open water;winter true,ssmicon low;open_water;1.0

Fleet Numerical Meteorology and Oceanography Center (FNMOC). To improve the run-time performance of ARKTOS, we convert all such files into the PGM format in which each pixel represents the ice concentration value at that location, computable through a conversion algorithm. For each feature, we convert its centroid (in image-based Cartesian coordinates) to latitude–longitude coordinates and map these coordinates onto the Cartesian coordinates of the corresponding concentration GRIB map (by date). Finally, we extract the concentration value, an integer between 0 and 100.

On a coastal sea ice image, in order to prevent land pixels from being factored into the segmentation and feature extraction process, we superimpose a land mask so that land pixels are distinctively designated. The land masks currently in use were provided to us by the Naval Research Laboratory, in NSIPS format, and are geocoded for easy mapping. To fuse land masks into ARKTOS, we created an attribute called *adjacent_to_land* and encoded six expert rules that involve the attribute. Any feature neighboring a land region will thus have its *adjacent_to_land* attribute set to true. In this manner, we are able to maintain a consistent use of land masks in the fusion process. These data have proven to be important in correctly identifying fast ice.

The NIC ice climatology dataset is a statistical computation describing extent and coverage of sea ice in specified areas of the Arctic and Antarctic Oceans. The data were compiled from 19 years of NIC Arctic and Antarctic sea ice analyses (covering most of the years 1972–1994) and are important for sea ice classification, as experts can draw inferences from the expected ice conditions in a region.

The current version of ARKTOS requires two sets of ice climatology data. The first set is the OCC, a map of the Arctic region for a particular month, where each pixel on the map represents the probability of that region containing ice in a 19-year span. Thus, the pixel value is between 0 and 19, with some other default values for land and null values. The second set is the MEDCT data, a map of the Arctic region for a particular month, where each pixel on the map represents the median concentration of ice in that region in a 19-year span. Thus, the pixel value is between 0 and 100, with some other default values for land and null values.

For each feature, we convert its centroid (in Cartesian coordinates) to latitude–longitude coordinates, map those coordinates onto the Cartesian coordinates of the corresponding climatology map, and extract the number of years of ice coverage value from the OCC files and the value of total ice concentration from the

MEDCT files. Finally, ARKTOS converts the above two measurements into facts, using the thresholds shown in Table IV for the symbol descriptors numyr and medct.

C. Rule-Based Classification

The rules of ARKTOS encode the knowledge extracted from expert analysts through a series of interviews and subsequently refined through testing and evaluation of prototype systems (the knowledge acquisition and refinements processes are detailed in [14]).

ARKTOS rules consist of a condition, a possible classification, and a weight. The condition is expressed as a collection of attribute–value pairs, where all attributes and their possible values are described in Tables III and IV. An example condition is

$$(\text{return} = \text{bright}) \, \text{AND} \, (\text{winter} = \text{true}) \, \text{AND} \, (\text{rounded} = \text{true}).$$

The resulting ice classification is one of four possible ice classes: open water, new ice, first-year ice, or multiyear ice. The weight of a rule indicates the mass of evidence in a classification, given the feature description defined in the condition part. A weight ranges between 0.1 and 1 for positive belief in a classification, and between -1 and -0.1 for negative belief. Weights of 1.0 and -1.0 indicate absolute certainty that a feature belongs or does not belong to a class, respectively. Currently, we have about 100 rules in our rule base, specifically for analyzing SAR sea ice images in the Beaufort Sea area.

For each feature, the rule-based classification module matches the facts associated with that feature with every rule in the rule base. If the conditions of a rule are matched, then the rule asserts a classification with a confidence value. For each feature, there are many rules that may be matched, and we may have rules supporting the feature as belonging to any number of classes. Hence, we need to combine these assertions in a consistent manner and determine which class (if any) is supported by the strongest evidence. To do so we use the Dempster–Shafer belief system [23].

Suppose that all of the ice classes that ARKTOS knows are of the frame of discernment or universe U. Thus, the set of all propositions (of all possible classifications) is $P(U)$, the power set of U. Let $m : P(U) \rightarrow [0, 1]$ be a function—a basic probability assignment. Let an assertion in favor of a classification be Γ. Then, the basic probability assignment function satisfies the conditions for a certainly false classification $m(\varnothing) = 0$
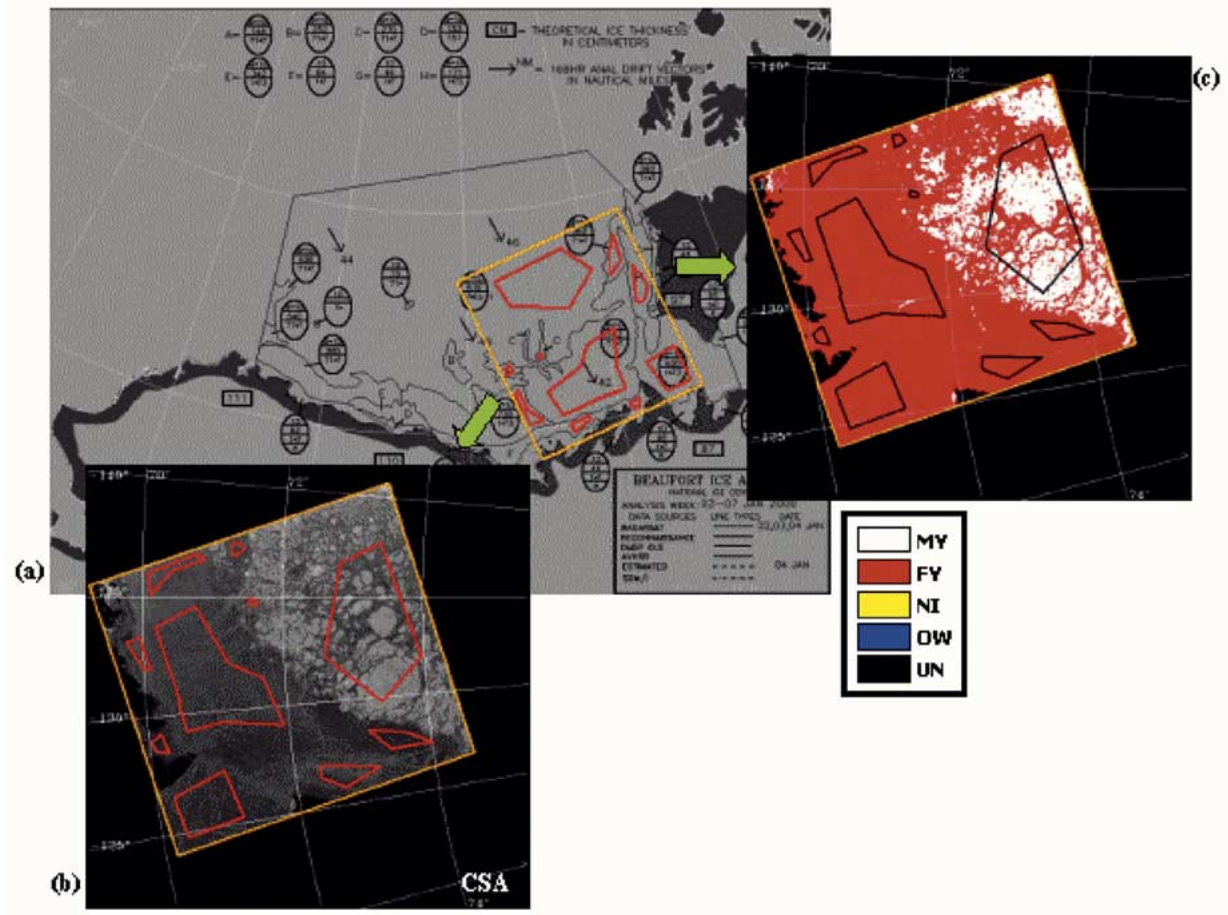
Fig. 3.   Polygons derived from (a) NIC ice charts, (b) same polygons overlain on the source RADARSAT image, and (c) ARKTOS classification result. (c) CSA 2000.

and for a certainly true classification $\sum_{\Gamma \subseteq U} m(\Gamma) = 1$. The belief function Bel : $P(U) \rightarrow [0,1]$ is defined in terms of the basic probability assignment m : $\mathrm{Bel}(\Gamma) = \sum_{\alpha \subseteq \Gamma} m(\alpha)$. This tells us the degree of belief associated with the classification Γ as the probability mass associated with Γ and its subsets. The plausibility of a classification is further defined as $\mathrm{Pls}(\Gamma) = 1 - \mathrm{Bel}(\neg\Gamma)$. Hence, a classification is always bound by [Bel, Pls] in terms of the confidence in its perceived truthfulness. To combine various pieces of evidence for building up beliefs in favor of various classifications, Dempster's rule of combination is used. Suppose we are given two assignments (two pieces of evidence), m_1 and m_2, and we want to combine them into a single piece of evidence. Hence, we compute

$$[m_1 \oplus m_2](\Gamma) = \frac{\sum_{\alpha \cap \beta = \Gamma} m_1(\alpha) m_2(\beta)}{1 - \sum_{\alpha \cap \beta = \varnothing} m_1(\alpha) m_2(\beta)}$$

where $\Gamma \neq \varnothing$, and $[m_1 \oplus m_2](\varnothing) = 0$.

For example, suppose that after matching the facts of a feature to our rule base we arrive at two assertions: old_ice with confidence 0.7 and open_water with confidence 0.2. Hence, m_1 corresponds to the mass supporting the feature to be {old_ice} = 0.7 and to be any of $\Theta = 0.3$, where Θ is the set of all ice classes; and m_2 corresponds to the mass supporting the feature to be {open_water} = 0.2 and to be any of $\Theta = 0.8$. Then,

TABLE VII
MEAN AND MEDIAN ABSOLUTE DIFFERENCES IN ICE CONCENTRATION FOR ALL *POLYGONS* BETWEEN ARKTOS AND NIC ICE CHART CLASSIFICATION

	Mean Absolute Difference	Median Absolute Difference
Total Ice	14.01%	9.96%
Multi-Year Ice	24.67%	13.53%
First Year Ice	28.75%	19.64%
New Ice	10.47%	0%
Open water	14.52%	10%

we can compute their combination m_3 using the rule of combination above, resulting in the classifications shown in Table V. Table V says that the evidence for the feature to be of the old_ice type is now 0.56; of the open_water type is now 0.06; and of one of the ice classes is now 0.24.

In this manner, all new evidence is incorporated into the previously accumulated beliefs consistently. From m_3, we can further compute the evidential interval [Bel, PLs] for each of the ice classes.

ARKTOS uses a modified Dempster–Shafer belief system to deal with the intricacies of sea ice classification. The modifications of the theory are as follows.

1) In ARKTOS, the classification of a sea ice feature into a set such as {open_water, old_ice}, Θ, or { } is not useful. Thus, the mass or evidence accumulated for such propositions is purged. For example, in Table V, after purging, the reweighted evidence for {open_water} is 0.09, and that

for {old_ice} is 0.85. Note that the Dempster–Shafer belief system assigns beliefs to all possible combinations of classifications, such as {open_water, old_ice}. That is, the system could indeed classify an object as open_water or old_ice with a degree of belief. However, we deem such a "hybrid" classification useless. Thus, we use the above purging technique, removing such ambiguity from our final classification. The purging process is simply removing the weights of these hybrid classifications, and renormalizing the remaining classifications so that they sum up to 1.0.

2) The original Dempster–Shafer belief theory does not include negative beliefs, something important in ARKTOS, since sea ice experts express knowledge of both positive and negative classification (e.g., if a particular feature exhibits certain attributes, then that feature cannot be of ice type "A"). Moreover, whenever such a negative assertion is made, it carries more weight than a corresponding positive assertion, as determined by sea ice experts. We modified the Dempster–Shafer belief system to account for negative assertions exactly the same way as for positive ones, and to weigh negative assertions more heavily.

3) We compute the product of belief and plausibility for each ice class for each feature and use that to determine the most credible ice class. If the product of the most credible ice class is below a certain threshold (0.25), then ARKTOS classifies the feature as "unknown."

4) Some rules have absolute certainty (positive or negative), i.e., they have a weight of 1.0 or −1.0. When such rules are fired, we deal with the classification differently. If a feature has competing classifications weighted with an absolute belief (1.0) or the same classification weighted with absolute belief (1.0) and absolute disbelief (−1.0), ARKTOS classifies the feature as "unknown." Otherwise, ARKTOS immediately classifies the feature according to the absolute beliefs, ignoring other "less than absolute" assertions.

Table VI lists a few rules whose antecedents are based on the attribute–value pairs, and the consequents are ice type classifications. To summarize, each rule has a weight value that indicates the contributing factor of the rule. This weight value is input as evidence or mass into the Dempster–Shafer belief system. The belief system combines the evidence (different weights) from different rules to obtain a belief measure for each ice class. It also computes the plausibility measure—a likelihood indicator—for each ice class based on the belief measure. Finally, ARKTOS multiplies the belief and plausibility measures to find the most credible ice class.

IV. EVALUATION OF ARKTOS

The evaluation of ARKTOS was performed using 54 RADARSAT images collected over the years 1998–2000. The locations of these images ranged over the Beaufort, Bering, and Chukchi regions, and the images were evenly distributed throughout the course of the year. None of the imagery was used previously in any of the ARKTOS fine-tuning efforts. Climatology coincident with the imagery was retrieved
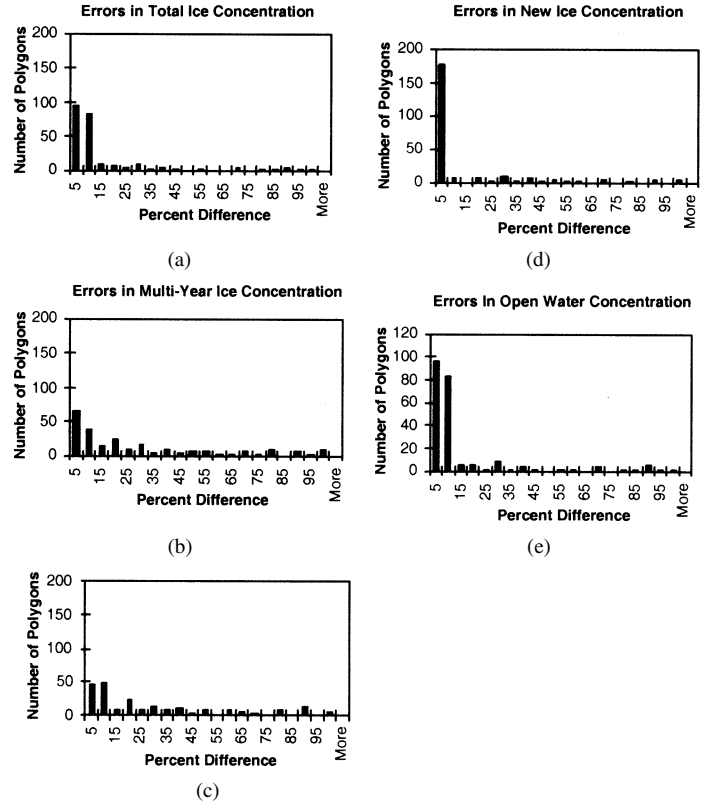

Fig. 4. Distributions of ice concentration differences (absolute value) for (a) total ice concentration, (b) multiyear ice, (c) first-year ice, (d) new ice, and (e) open water.

TABLE VIII
MEAN AND MEDIAN ABSOLUTE DIFFERENCES IN ICE CONCENTRATION FOR ALL *IMAGES* BETWEEN ARKTOS AND NIC ICE CHART CLASSIFICATION

	Mean Absolute Difference	Median Absolute Difference
Total Ice	8.37%	5.54%
Multi-Year Ice	22.81%	17.67%
First Year Ice	23.54%	15.08%
New Ice	4.31%	4.33%
Open water	8.55%	5.54%

from the NIC, and the Fleet Numerical and Meteorological Oceanographic Command (FNMOC) provided the necessary coincident ancillary SSM/I CAL/VAL ice concentration data [22].

The ARKTOS ice classification results were then compared to the total and partial ice concentrations obtained from coincident NIC ice charts. Using a combination of ARCINFO and NSIPS tools, polygons representing areas of common partial and total ice concentrations were derived from the NIC ice charts and then overlain onto the ARKTOS ice classification product (Fig. 3). The ARKTOS ice classification product consists of a five-value image with pixel values representing multiyear ice, first-year ice, new ice, open water, and unknown/undecided. Values obtained from a histogram of the ARKTOS ice classification product within an area defined by a polygon derived from the NIC ice chart were then used to calculate ARKTOS' estimates of total and partial ice concentration inside that polygon.

Approximately 100 NIC ice concentration charts were used as "truth" for the comparison. These weekly ice concentration
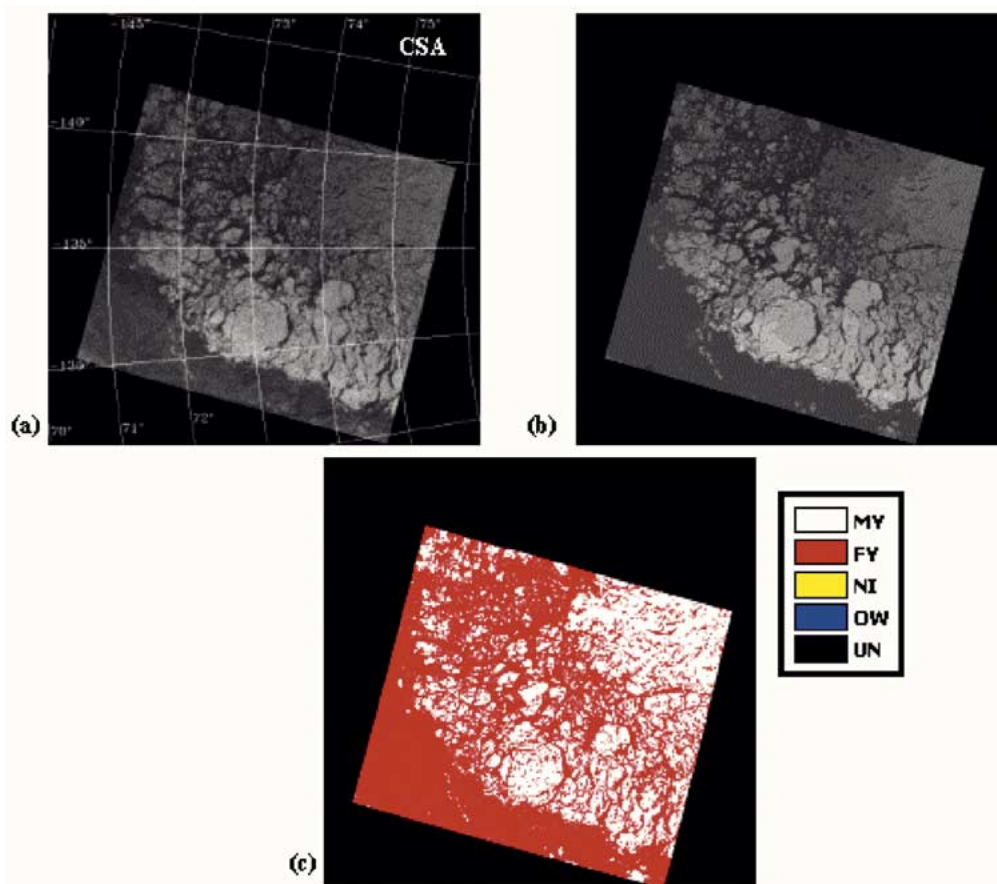
Fig. 5.   Typical ARKTOS results for a mixture of multiyear and first-year ice. (a) RADARSAT image from February 20, 2000, (b) ARKTOS segmentation, and (c) ARKTOS classification product. Note the accuracy of the segmentation (a) CSA 2000.

charts are produced by a team of experienced ice analysts from a combination of different sources of satellite imagery received at the NIC. When available, *in situ* observations are also used. Stringent classification criteria are used by the ice analysts in order to insure the accuracy and consistency of their product. Although the ice charts are not ground truth per se, they represent the absolute best guess about the types and concentrations of ice that can be inferred from the data. Note that by assuming the NIC ice charts as "truth," our study uses absolute difference between ARKTOS' classification and the ice charts as the classification error.

Although the NIC ice charts are produced by ice experts, they are still the product of a human interpretation of the data and, as such, are not entirely free of uncertainty. Total concentrations are generally represented by a range of values that can span up to 20%. For example, an area in an NIC ice chart of multiyear pack ice may be represented as having 80% to 100% total ice concentration, with no indication within the area where the concentration is 80% and where it is 100%. This value range, which is not constant from ice chart polygon to ice chart polygon, can introduce error when compared to the exact total ice concentrations calculated from the ARKTOS classifications. To minimize this error, NIC total ice concentration estimates used in the comparison were calculated from a sum of the partial ice concentrations in the polygons. There may also be some variability introduced

by the subjective nature of the NIC ice analyses themselves. Although the Beaufort region is routinely analyzed at the NIC by one of their best analysts, interpretation of the input data may vary from analyst to analyst. There will be variations among the analysts in the estimates of concentration boundaries and in the ice types within those boundaries. These interanalyst variations can be up to 10% for the estimate of total concentration and up to 20% for the estimate of partial concentration. This variability can also increase based on differences in the level of expertise among the analysts.

Care was taken to ensure that error was not introduced into the analysis when using the ice charts for comparison. Most of the imagery used in the evaluation went directly into the production of the ice charts, thus rendering the evaluations a direct comparison of ARKTOS' results to those of the ice analysts. For cases where the analyzed imagery was not used as input to the ice chart, care was taken to make sure that whatever product was used (Advanced Very High Resolution Radiometer (AVHRR), Operational Linescan System (OLS), SSM/I) was generated on the same day as the RADARSAT imagery. As a final check, the imagery was visually inspected to reaffirm that the ice chart concentrations and classifications were accurate at the feature scale of the imagery. In a few cases, visual inspection values did not mirror ice chart values. In some of these cases, the mismatch was due to the dynamic nature of the ice pack (the effects of ice
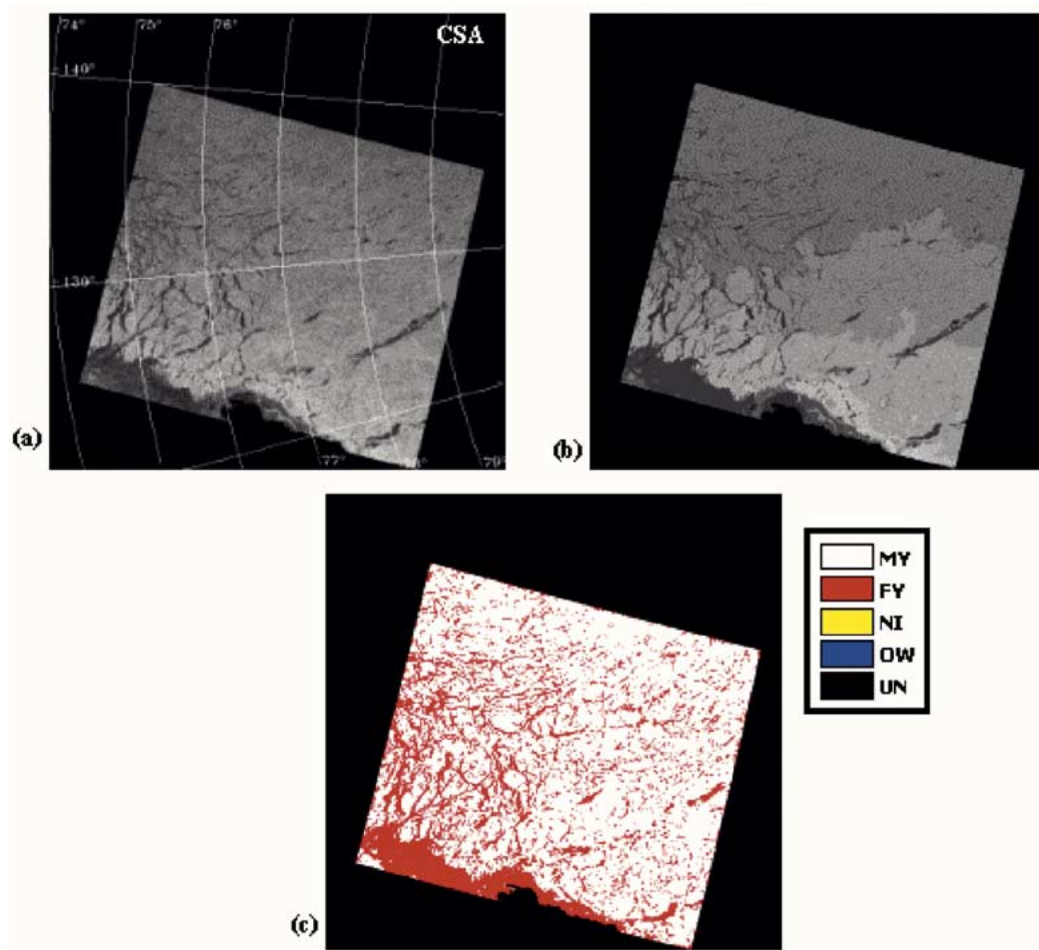
Fig. 6.   Typical ARKTOS results for multiyear ice pack in winter. (a) RADARSAT image from February 20, 2000, (b) ARKTOS segmentation, and (c) ARKTOS classification product. (a) CSA 2000.

motion). These cases were removed from the analysis. Other cases were sent to the NIC for a more in-depth analysis at the image feature level. Every effort was made to visually evaluate and validate the NIC ice concentration estimates at the image feature level prior to the comparison to the ARKTOS results.

ARKTOS' estimates of ice type and ice concentration within areas of analysis derived from NIC ice chart polygons were then compared to the matching NIC values within those areas. For every polygon analyzed, the difference between the total and partial NIC ice concentrations and the ARKTOS ice concentration estimates was calculated. These differences were examined with respect to latitude, longitude, and data and were listed as either underestimating or overestimating of ice concentration by ARKTOS. Statistics were then derived from all the pairs of values.

A few trends were observed in the values of the differences that were calculated. One of the most noticeable trends was the large number of overestimations of total ice concentration by ARKTOS (and paired underestimations of open water) at 5% and 10%. These values may reflect uncertainty in the ice concentration values in the ice charts rather than real error in ARKTOS. The majority of these errors are associated with regions of multiyear pack ice in the NIC ice charts. These areas are usually represented by a total ice concentration of 90% to

100%, with partial concentrations of 80% multiyear ice, 10% first-year ice, and a trace of new ice. Efforts were made to minimize error introduced by a range of total concentration values by comparing ARKTOS total ice concentration values to the sum of the NIC ice chart partial values. But there was still a question of how to represent the "traces" of new ice. A trace can represent values from 0% to 9%, making the sum of the partials equal to anywhere from 90% to 99%. When the RADARSAT images paired with these NIC values are examined, the areas that make up the consequent 0% to 9% of open water are not readily discernable by the ice analysts. They rely more heavily on clues from AVHRR and OLS data, or from meteorology and season to estimate this open water. As the ARKTOS classification is driven by the RADARSAT image, it cannot compensate for what is not present in the image, and it calls the entire region 100% ice, introducing a 5% to 10% overestimation of ice throughout.

There appear to be some seasonal trends in the errors. One is in the classification of new ice. With the exception of a few values on day 200 (mid-July), most of the large errors in estimation of new ice appear to occur between day 260 (mid-September) and day 350 (mid-December) and seem to be somewhat paired with errors in the estimation of first-year ice. When the imagery is analyzed, these errors appear to come from areas
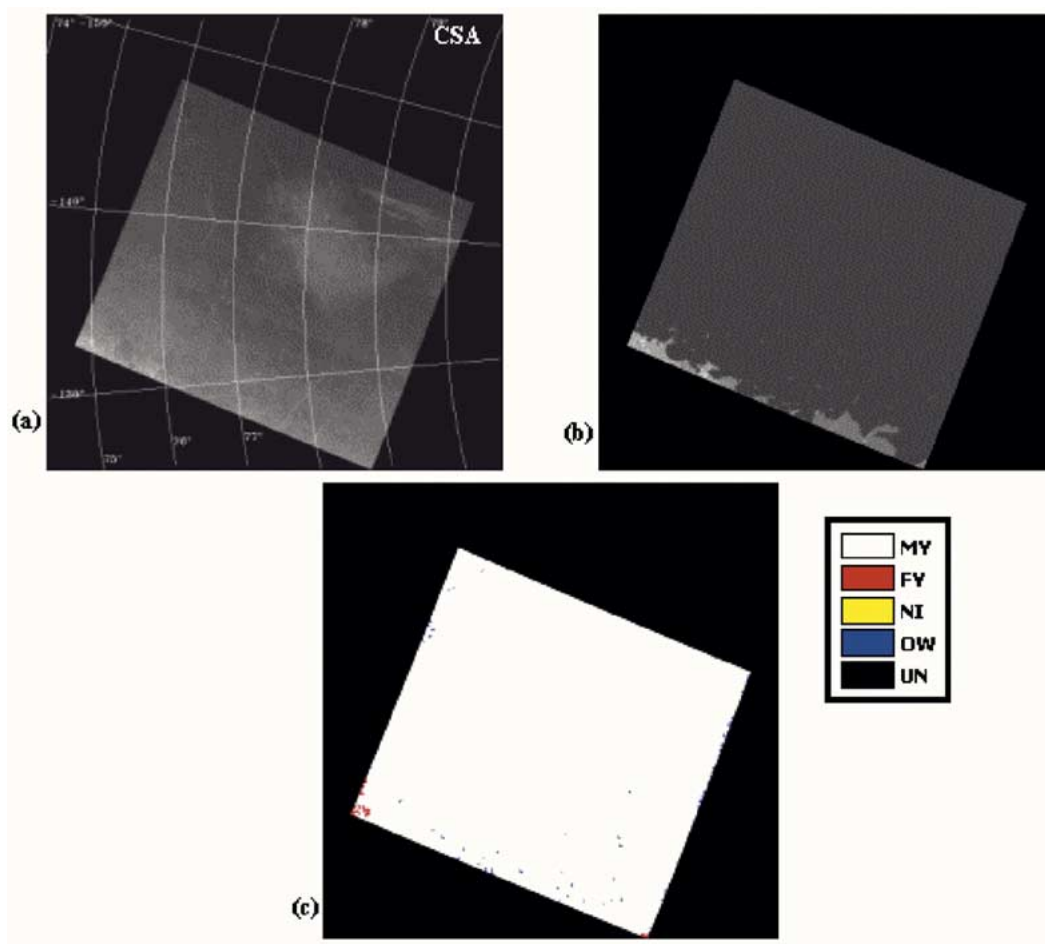
Fig. 7.   Typical ARKTOS results for multiyear ice pack in summer. (a) RADARSAT image from August 9, 2000, (b) ARKTOS segmentation, and (c) ARKTOS classification product. (a) CSA 2000.

classified by the NIC as young ice in the central Beaufort. For the purposes of this comparison, young ice was collected together with much thinner ice types into a new ice class, although it can be nearly as thick as FY ice. It is interesting to see that these errors are present mostly in the fall and early winter. It may be a result of the fact that polygons which cover areas that are considered to be young ice switch to first-year ice (in the NIC ice charts) each fall, when ice thickness increases as temperatures drop. The other trend observed is a slight increase in the size of the errors for the total ice concentration and the multiyear and first-year partial ice concentrations during the summer. Low image contrast and intermittent backscatter inversion (where water pixels appear brighter and ice pixels appear darker in the image, due to the wind-roughened water surface during the summer) play a large role in classification error during the summer.

With respect to latitude, there are no significant trends in the differences for the total ice concentration or the partial concentration of open water, other than the overestimation of ice (and underestimation of water) described previously. MY and FY partial ice concentration differences display a spread of overestimations and underestimations for both ice types beyond 70°, although there appears to be slightly more overestimations of FY ice (and coincident underestimations of MY ice) between 70° and 75°. Below 70° ARKTOS appears to be underestimating

FY ice and overestimating MY ice, although the differences are not completely correlated. The only area of the analysis with latitudes less than 70° was the Bering Sea, so these values may represent some confusion by ARKTOS among MY, FY, and open water in the marginal ice zone. The largest errors in the estimation of new ice occur between about 68° and 75°. Examination of these areas indicates that they are from polygons located in the central Beaufort and are correlated with large errors in the FY ice estimates, so this set of results is similarly indicative of young ice/FY ice confusion.

There appear to be no significant trends in the differences between ARKTOS and NIC concentration for MY and FY ice with respect to longitude, but total ice concentration, open water (which is paired with total concentration), and new ice values show larger errors at specific longitudes. In the case of the total ice concentration values and the open water values, errors are on the order of 5% to 10% for longitudes between −120° and −160°, and range fairly widely between −160° and −180°. These trends are mostly due to the types of ice features at the different locations. The central Beaufort lies between −120 and −160°, and this area tends to be filled with the standard MY/FY ice pack described above, with the errors as described above. Between −160° and −180°, the Beaufort is merging with the Chukchi and Bering Seas. These areas are very dynamic, and as such, have a wider range of mixtures of open water and
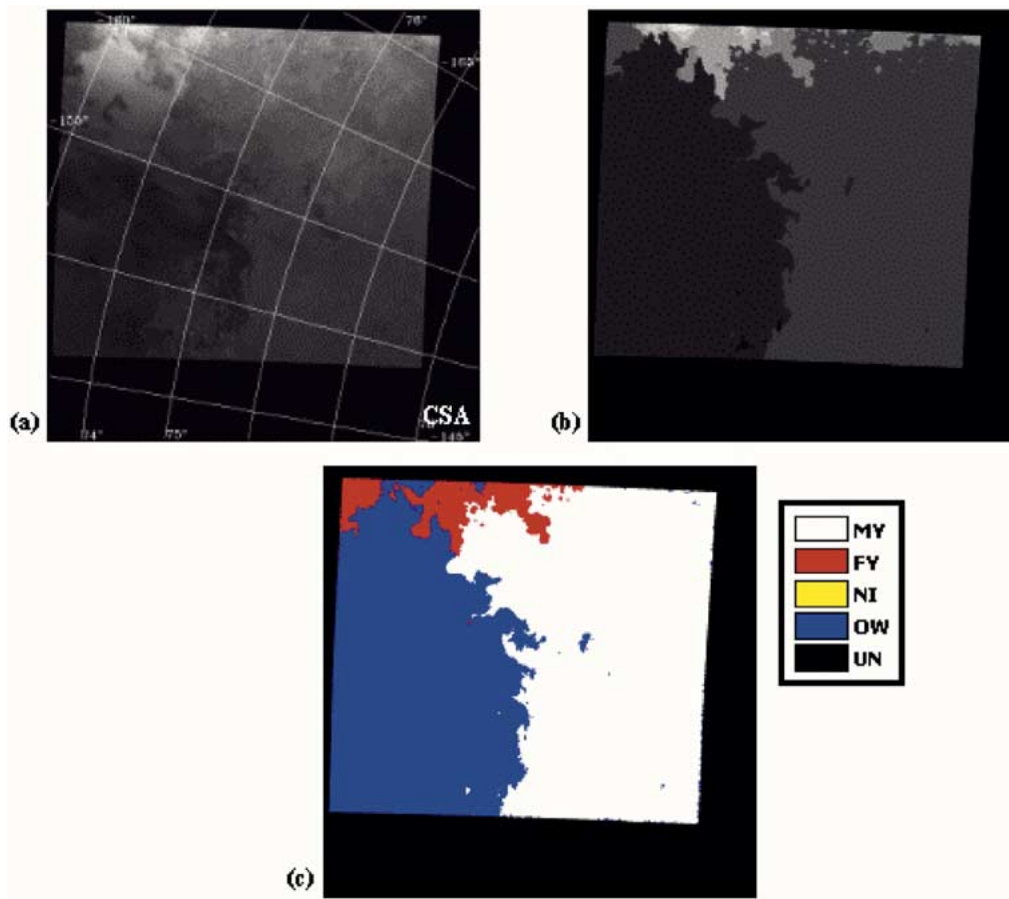
Fig. 8.   Typical ARKTOS results for the ice edge in summer. (a) RADARSAT image from August 8, 1998, (b) ARKTOS segmentation, and (c) ARKTOS classification product. (a) CSA 1998.

ice. The cluster of ARKTOS underestimations of new ice between $-140°$ and $-160°$ is caused primarily by confusion on ARKTOS' part between young ice and first-year ice.

Table VII presents the mean and median absolute differences between the total and partial ice concentration estimates derived by ARKTOS for each polygon and those derived from the NIC ice charts. Fig. 4 shows distributions of these absolute differences. Although the mean absolute differences are somewhat large, the median values and histograms show that most errors for total concentration (and hence, also open water) are on the order of 5% to 10%. Those for the partial ice concentrations (MY, FY, and NI) are on the order of 5% to 20%. These values coincide with the expected variability in the ice charts of 10% for total ice concentration and 20% for partial ice concentration, and on par with results from other classification algorithms [24].

These measurements provide a sense of ARKTOS error for a random set of comparisons, but they are somewhat biased toward the more complex images in the dataset. Complex images, having a wider range of combinations of different ice types and concentrations, contribute more polygons per image to the mean and median estimates, but their area does not necessarily represent a large spatial portion of the Beaufort. In order to provide a more spatially based examination, an analysis of ARKTOS accuracy on a per-image basis was also performed, with polygon size being taken into account. The results of this analysis are presented in Table VIII. When the results from each image are allowed to contribute equally to the error estimates, the mean absolute error for the total ice concentration is 8.4%, while that for the partial ice types ranges from 4.3% to 23.5%. Median values are 5.5% for the total ice concentration and 4.3% to 17.7% for the partial ice types.

Results taken over the course of this evaluation indicate that the inclusion of SSM/I ice concentration data is necessary for the optimal operation of ARKTOS. SSM/I data should be made available to ARKTOS whenever possible. SSM/I data aid ARKTOS in the delineation of open water and ice and assist in the identification of ice along the ice edge. When operated without the inclusion of SSM/I input, ARKTOS' accuracy varies greatly. For images consisting of 100% total ice concentration throughout, the results of operating ARKTOS without SSM/I input are almost identical to those results obtained when SSM/I input is included. Results obtained over areas of open water or along the ice edge can be correct when SSM/I data are omitted, but are usually flawed. Differences in the source of the SSM/I ice concentrations will also cause differences in the results obtained from ARKTOS. Currently, threshold values for the SSM/I ice concentration feature attribute (SSMICON) are geared toward the range of values obtained from the FNMOC CAL/VAL ice concentration algorithm. If other sources of SSM/I ice concentration values are used, the threshold file should be modified accordingly before operation.
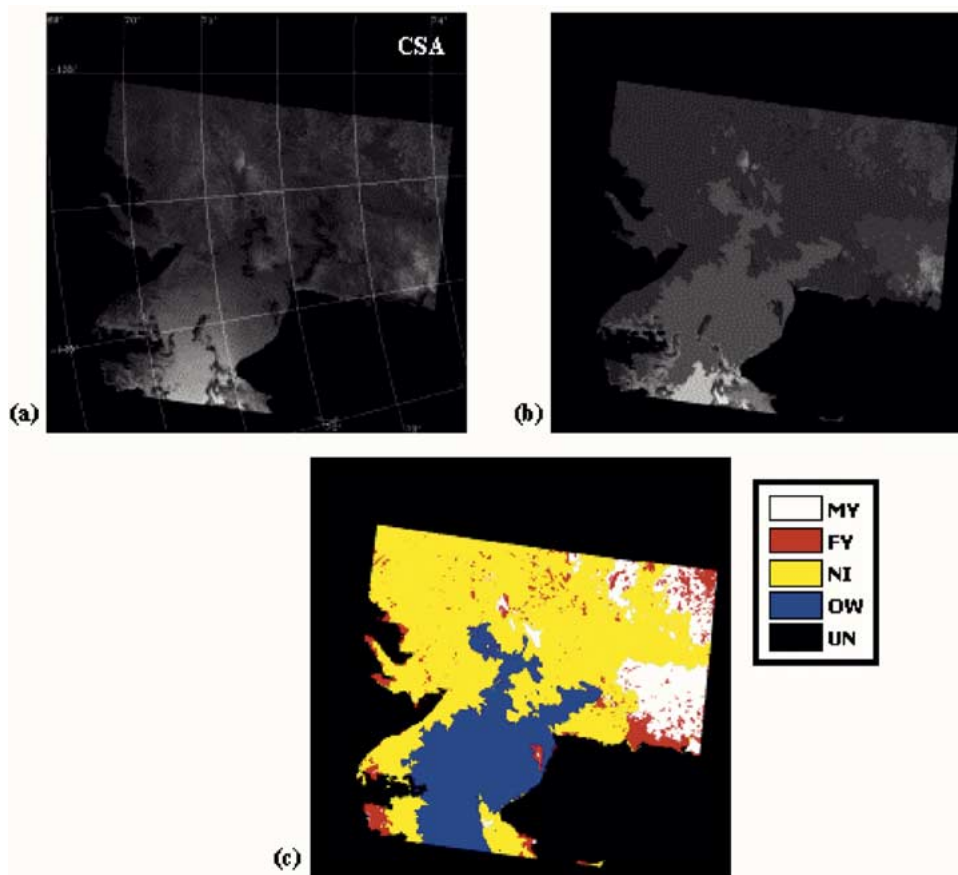
Fig. 9. ARKTOS capturing regions of new ice and pockets of multiyear ice within the new ice. (a) RADARSAT image from October 31, 1999, (b) ARKTOS segmentation, and (c) ARKTOS classification product. Note: the black areas in (c) within the sea ice map are land. (a) CSA 1999.

Figs. 5–9 represent results typical of ARKTOS performance. Fig. 5 presents the results for an image of the MY/FY ice pack in the central Beaufort taken on February 20, 2000. ARKTOS performs an excellent job of separating first-year ice areas from multiyear ice areas and of classifying them correctly. Within the multiyear ice pack, it can identify and segment out some individual floes, but also aggregates some floes into larger features that no longer obey the standard set of rules the ice analysts have developed for multiyear ice floes. Even with the change in scale and appearance brought about by this aggregate segmentation, ARKTOS is able to classify the multiyear pack ice correctly. Figs. 6 and 7 present ARKTOS results for winter (February 20, 2000; Fig. 6) and summer (August 9, 1998; Fig. 7) images of ice pack. In the winter image, the multiyear ice pack is broken down into four main features roughly based on image intensity. In the summer image, low image contrast causes ARKTOS to segment the image into basically one large feature. In both cases, ARKTOS correctly classifies the multiyear pack and the surrounding first-year ice areas.

Fig. 8 provides an example of ARKTOS performance in the marginal ice zone (MIZ). The image is from August 7, 1998 in the western Beaufort. Most of the image is correctly segmented, except for a small area of ice surrounded by open water in the eastern half of the image. ARKTOS classifies all but the very western edge of the image correctly. It calls this area first-year ice, while the NIC ice charts indicate that it is really an area of low-concentration multiyear ice. This error illustrates

an interesting conundrum within the operational capability of ARKTOS. ARKTOS does an excellent job of separating this area of low ice concentration from adjacent areas of high ice concentration and open water, but does not have the capability of labeling it as a low ice concentration area. No matter what ARKTOS calls this area, it will be at 100% concentration. Segmentation parameters can be tweaked so that ARKTOS can pick out the individual floes within this area in order to achieve an overall lower ice concentration, but such tweaking would cause major segmentation errors elsewhere.

Fig. 9 presents a complex example of the ARKTOS algorithm at work. The raw RADARSAT image, taken on October 31, 1999, is of the eastern Beaufort during freeze-up. ARKTOS accurately delineates darker open water areas from darker new ice areas, and brighter new ice areas from multiyear ice. In addition, it picks out individual features within the new ice. ARKTOS correctly classifies both brighter and darker open water areas, which it has segmented into separate features, as open water. It also correctly classifies an area of darker multiyear ice in the northwest corner of the image as multiyear ice. Both the ARKTOS result and the NIC ice chart display an ice pack made up of predominantly new ice on the ice edge and in the south, merging into a mixture of new ice and multiyear ice toward the northwest. Although the ARKTOS classification does not match the NIC ice charts exactly pixel for pixel everywhere, it is a fairly accurate representation of the ice and water present overall.

V. Discussion

We have presented the ARKTOS intelligent system for satellite sea ice image analysis. ARKTOS is a fully automated system that classifies sea ice images by mimicking the reasoning process of sea ice experts. The approach is a feature-based, rule-based classification system supported by multisource data fusion. ARKTOS underwent a rigorous evaluation process against actual operational sea ice charts generated and maintained by the U.S. National Ice Center and achieved good results in high-Arctic areas, even when compared to human expert classification. ARKTOS has a flexible and extensible design due to the modularity of the knowledge bases and the suite of GUI-based software for effective evaluation and refinement. The NIC has also operationalized ARKTOS and the NIC analysts use the ARKTOS output as one of the information sources they study to develop their standard operational map product.

The major contribution of ARKTOS is that it proves that it is possible to develop a fully automated, accurate intelligent classifier of natural scenes that starts from the original data sources and, without any human intervention, produces a final classification product in near real-time. ARKTOS is also the first fully automated, near real-time, operational sea ice classifier, and it has been shown to achieve very good classification results compared to these of human experts, with mean absolute difference for the total ice concentration 8.4%, and that for the partial ice types ranging from 4.3% to 23.5%. Finally, the methodologies we have employed, the lessons we have learned during our research, and the innovative approaches to image processing, data manipulation, and knowledge engineering during the past ten years of developing, prototyping, refining, and testing ARKTOS provide a detailed and successful roadmap for building other intelligent geophysical classification systems.

The following are several possible modifications to ARKTOS that may further improve its performance:

1) image correction during the preprocessing stage to compensate for nadir ambiguities and near-range brightness due to $R^{**}4$ fall-off;
2) adaptive segmentation using spatial and temporal information;
3) consideration of a previous week's ice analysis to account for the persistence and continuity of the ice;
4) modification of ARKTOS for use on ENVISAT/RADARSAT-2 polarimetric data, since HV polarization data should improve ARKTOS' ability to discriminate between ice and open water;
5) consideration of AVHRR and OLS data as part of the ancillary data input for ARKTOS, since these data are often used by ice analysts to delineate between ice and open water;
6) consideration of weather and other environmental parameters, since these factors are often taken into account by ice analysts to help delineate new ice from first-year ice;
7) further development of shape and texture attributes;
8) refinement of ARKTOS' knowledge bases and expansion to cover the entire Arctic region.

References

[1] C. Tsatsoulis and R. Kwok, Eds., *Analysis of SAR Data of the Polar Oceans*. Berlin, Germany: Springer-Verlag, 1998.

[2] T. Matsuyama and V. Hwang, "SIGMA, a knowledge-based aerial image understanding system," in *Advances in Computer Vision and Machine Intelligence*. New York: Plenum, 1990.

[3] F.-J. Wang, "A knowledge-based vision system for detecting land changes at urban fringes," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 136–145, Jan. 1993.

[4] J. Ton, J. Sticklen, and A. K. Jain, "Knowledge-based segmentation of Landsat images," *IEEE Trans. Geosci. Remote Sensing*, vol. 29, pp. 222–232, Mar. 1991.

[5] P. G. Foschi and D. K. Smith, "Detecting subpixel woody vegetation in digital imagery using two artificial intelligence approaches," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 5, pp. 493–500, 1997.

[6] D. S. Kimes, P. R. Harrison, and P. A. Ratclife, "A knowledge-based expert system for inferring vegetation characteristics," *Int. J. Remote Sens.*, vol. 12, pp. 1987–2020, 1991.

[7] A. Srinivasan and J. A. Richards, "Knowledge-based techniques for multi-source classification," *Int. J. Remote Sens.*, vol. 11, pp. 505–526, 1990.

[8] F. Fetterer, B. Cheryl, and J. P. Ye, "Multi-year ice concentration from Radarsat," in *Proc. IGARSS*, Singapore, 1997.

[9] R. Kwok, "The RADARSAT geophysical processor system," in *Analysis of SAR Data of the Polar Oceans*, C. Tsatsoulis and R. Kwok, Eds. Berlin, Germany: Springer-Verlag, 1998, pp. 235–279.

[10] L.-K. Soh, C. Tsatsoulis, T. Bowers, and A. Williams, "Representing sea ice knowledge in Dempster–Shafer belief system," in *Proc. IGARSS*, Seattle, WA, 1998, pp. 2234–2236.

[11] L.-K. Soh and C. Tsatsoulis, "Multisource data and knowledge fusion for intelligent SAR sea ice classification," in *Proc. IGARSS*, Hamburg, Germany, 1999, pp. 68–70.

[12] C. Tsatsoulis, L.-K. Soh, C. Bertoia, and K. Partington, "Intelligent fusion of multisource data for sea ice classification," in *Proc. ACAI'99 Workshop on Intelligent Techniques for Spatio-Temporal Data Analysis in Environmental Applications*, Chania, Greece, July 5–16, 1999.

[13] L.-K. Soh and C. Tsatsoulis, "ARKTOS: A knowledge engineering software package for satellite sea ice," in *Proc. IGARSS*, Honolulu, HI, 2000, pp. 696–698.

[14] ——, "ARKTOS: A knowledge engineering software tool for images," *Int. J. Human-Comput. Stud.*, vol. 57, pp. 469–496, 2003.

[15] C. Bertoia, D. Gineris, K. Partington, L.-K. Soh, and C. Tsatsoulis, "Transition from research to operations: ARKTOS: A knowledge-based sea ice classification system," in *Proc. IGARSS*, Hamburg, Germany, 1999, pp. 1573–1574.

[16] F. Fetterer, D. Gineris, and C. Johnson, "Remote sensing aids in sea-ice analysis," *EOS Trans. Amer. Geophys. Union*, vol. 74, no. 24, pp. 265–268, 1993.

[17] K. R. Dedrick, K. Partington, M. Van Woert, C. A. Bertoia, and D. Benner, "U.S. National/Naval Ice Center digital sea ice data and climatology," *Can. J. Remote Sens.*, vol. 27, no. 5, pp. 457–475.

[18] D. Haverkamp, L.-K. Soh, and C. Tsatsoulis, "A comprehensive, automated approach to determining sea ice thickness from SAR data," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 46–57, Jan. 1995.

[19] L.-K. Soh and C. Tsatsoulis, "Unsupervised segmentation of ERS and RADARSAT sea ice images using multiresolution peak detection and aggregated population equalization," *Int. J. Remote Sens.*, vol. 20, no. 15–16, pp. 3087–3109, 1999.

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

[21] A. I. Watson, "A new method of classification for Landsat data using the watershed algorithm," *Patt. Recognit. Lett.*, vol. 6, pp. 15–19, 1987.

[22] J. P. Hollinger, "DMSP Special Sensor Microwave/Imager calibration/validation final report," Naval Res. Lab., Washington, DC, vol. 2, 1989.

[23] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.

[24] D. J. Gineris and F. M. Fetterer, "The Joint Ice Center SAR Workstation: Algorithm evaluation," Naval. Res. Lab., Washington, DC, NRL Memo. Rep. 7019, 1993.

Costas Tsatsoulis (M'88–SM'98) received the Ph.D. degree from Purdue University, West Lafayette, IN, in 1987.

He is currently a Professor of electrical engineering and computer science and is Director of the Intelligent Systems and Information Management Laboratory, University of Kansas, Lawrence. His research interests are in artificial intelligence, multiagent systems, and intelligent image processing. He has over 100 publications in these areas and is coeditor of *Analysis of SAR Data of the Polar Oceans* (Berlin, Germany: Springer-Verlag, 1998).

Dr. Tsatsoulis is a member of the ACM and the AAAI.

Denise Gineris, photograph and biography not available at the time of publication.

Leen-Kiat Soh (S'91–M'98) was born in Johor, Malaysia. He received the B.S. and M.S. degrees and the Ph.D. degree (with honors) in electrical engineering from the University of Kansas, Lawrence, in 1991, 1993, and 1998, respectively.

He is currently an Assistant Professor in the Computer Science and Engineering Department, University of Nebraska, Lincoln. His research interests include intelligent image analysis, multiagent systems, machine learning, and computer-aided education systems.

Dr. Soh is a member of the ACM, AAAI, Phi Kappa Nu, Tau Beta Pi, Eta Kappa Nu, and Phi Beta Delta.

Cheryl Bertoia received the B.A. degree in geography from the University of Rhode Island, Providence, and the M.S. degree in computer systems from the University of Maryland, College Park, in 1983 and 1997, respectively.

She has worked in operational sea ice mapping and related research activities at the U.S. National Ice Center (formerly the Navy/National Oceanic and Atmospheric Administration Joint Ice Center), Washington, DC since 1985. In recent years, she has concentrated on validation and operational implementation of sea ice classification algorithms for synthetic aperture radar and active and passive microwave imagery.