

# Multisource Data and Knowledge Fusion for Intelligent SAR Sea Ice Classification

Leen-Kiat Soh and Costas Tsatsoulis

Department of Electrical Engineering and Computer Science

University of Kansas, Lawrence, Kansas, USA

ITTC, 2291 Irving Hill Road, Lawrence, KS 66045 USA

Tel: (785) 864-7764 Fax: (785) 864-0387 E-mail: {lksoh,tsatsoul}@itcc.ukans.edu

**Abstract** – In this paper we describe the fusion of various data and knowledge sources for intelligent SAR sea ice classification, thereby addressing the weaknesses of each information source while improving the overall reasoning power of the classifier. We equip our ice classification system, ARKTOS, with the capability of analyzing and classifying images unsupervised by emulating how a human geophysicist or photo-interpreter classifies SAR images. To imitate human visual inspection of raw images, we have designed and implemented a data mining application that first categorizes pixels into regions, and then extracts for each region a complex feature set of more than 30 attributes. In addition, we have incorporated other sea ice data and knowledge products such as ice concentration maps, operational ice charts, and land masks. Finally, we solicited human sea ice expertise as classification rules through interviews, and collaborative refinements during the early-stage evaluations. Using a Dempster-Shafer belief system, we are able to perform multisource data and knowledge fusion in ARKTOS' rule-based classification. ARKTOS has been installed at the National Ice Center and Canadian Ice Service.

## INTRODUCTION

Our ARKTOS (Advanced Reasoning using Knowledge for Typing Of Sea ice) software performs automated, intelligent SAR sea ice classification. Once it is deployed, it will be incorporated into the flow of the operations to (1) assist geophysicists or photo-interpreters in identifying sea ice classes, and (2) perform pre-processing tasks to help streamline the workflow. To equip ARKTOS with the capability of analyzing and classifying images unsupervised, we aim at emulating how humans classify SAR sea ice images. We realize that humans use expert knowledge that they have learned and accumulated through experience. In addition, humans integrate other data and knowledge sources into their classification process for support. Therefore, in the ARKTOS project, we have had to address the following issues. First, in order to capture the expert knowledge, ARKTOS needs to provide the humans visual cues that can be quantified. Second, to imitate human reasoning process, ARKTOS needs to consider other viable sea ice data and knowledge sources as well. As a result, our ARKTOS project has involved data mining and multisource data and knowledge fusion.

In this paper, we describe the data mining process that transforms original images into a database of features with numerous attributes and facts. Then, we show how we incorporate the ancillary data and other data and knowledge sources into the system. In addition, we briefly discuss the knowledge transfer work from sea ice geophysicists and photo-interpreters to quantifiable attributes and rules. Finally, we use a Dempster-Shafer belief system to fuse all pieces of information for intelligent SAR sea ice classification. Fig. 1 shows the multisource data and knowledge fusion of ARKTOS for intelligent SAR sea ice classification. Even though not explicitly shown in the figure, note that human expert knowledge has been utilized to define and refine the data mining process that generates attributes and symbolic facts.

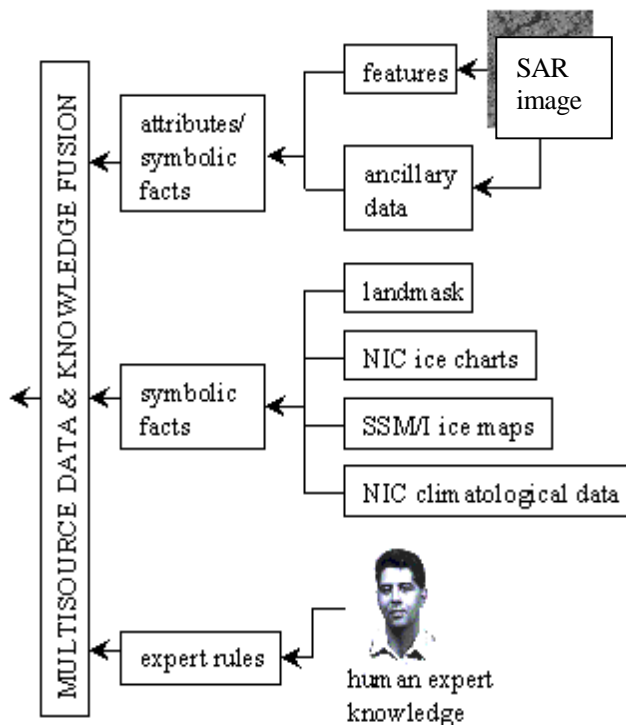


Fig.1 The various data and knowledge sources used incorporated in ARKTOS.

## DATA MINING

The first data source is the raw, original images. To imitate human visual inspection of raw, original images, first we perform data mining to extract features from the images.

Each feature will then be *visualized* computationally using a complex knowledge feature set of more than 30 attributes of intrinsic, geometric, positional, and comparative types. Attributes such as area, contrast, intensity, perimeter, roughness, orientation, elongation, roundness, irregularity, etc. are then further combined with expert knowledge to introduce symbolic facts such as *size=large*, *blob=true*, *smoother=true*, etc. which can be ingested by the rule-base system of ARKTOS. Hence, this data mining phase extracts from raw images a database of symbolic facts—a database in which other pieces of information are also represented.

## MULTISOURCE DATA AND KNOWLEDGE

To complement the information mined from the raw image, we consider other information sources: (1) ancillary data, (2) land masks, (3) NIC ice charts, (4) SSM/I concentration maps, and (5) climatology data.

### Ancillary Data

This data comes with the raw image as the header or leader information that specifies the date when and the geo-coordinates where the image was taken. This data is integrated into the database through several symbolic facts: *summer=true*, *winter=true*, *west\_of=xxxx*, *east\_of=xxxx*, and *latitude<=xxxx*, where *xxxx* is a coordinate value in degrees. Each SAR processing facility generally has a different ancillary data format to accompany its images. Thus, ARKTOS was designed to manage the differences. In addition, the ancillary data is also used to establish inter-source information coordination—to select corresponding ice chart and SSM/I concentration map.

### Land Masks

The land masks are used to identify land pixels. On a coastal sea ice image, in order to prevent land pixels from being factored into the data mining process (segmentation and feature extraction), we superimpose a land mask so that land pixels are distinctively designated. The land masks currently in use are provided to us by the Naval Research Laboratory and are geocoded to provide easy mapping. The knowledge of land is a vital piece of evidence used in human classification of sea ice. Thus, to fuse land masks into ARKTOS, we have created an attribute called *adjacent\_to\_land* and have 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.

### NIC Ice Charts

During the first week of September 1996, the National Ice Center (NIC) began routinely producing regional-scale analyses for all sea ice covered seas located (clockwise) between the longitudes of 95E and 95W in the northern hemisphere. This global area, known as the "West Arctic" to

the NIC, is comprised of fifteen analysis regions. Analyses of current ice conditions are produced weekly for regions in which sea ice is present.

All NIC regional sea ice analyses are derived from the near real-time integration of remotely sensed and in-situ oceanographic/meteorological observations. These analyses are produced following standard analysis procedures which optimize the use of data that vary widely in availability, scale, and resolution, from various operational data sources: satellite derived data, aerial ice reconnaissance, ship/shore station observations, drifting buoy reports, meteorological guidance products, ice prediction model output, climatology and sea ice information obtained from international partners such as foreign ice services. Sea ice experts or photo-interpreters often refer to NIC ice charts during sea ice classification due to its coverage and availability.

Our ice chart database manager has been designed and implemented to obtain the date and location information from the ancillary data to (1) locate the correct ice chart (past nearest date) from its collection, and (2) locate the ice code value (geo-coordinates) following the specifications in the World Meteorological Organization (WMO) SIGRID format for sea ice. It then translates the value to a piece of information such as the following:

[ *total concentration : 4/10-6/10*

[ *concentration for Old Ice : 5/10 with form of Undetermined/Unknown*

[ *concentration for Thick First Year Ice : Less than 1/10 (open water) with form of Undetermined/Unknown*

The above information will then be incorporated into the reasoning process of ARKTOS. Currently, the ice chart database manager has been taken out of ARKTOS' operation, awaiting a revised NIC ice chart format soon to be released.

### SSM/I Ice Concentration Maps

ARKTOS has the capability of analyzing images using DMSP SSM/I daily polar gridded sea ice concentration maps generated with the NASA Team algorithm, produced and distributed by the National Snow and Ice Data Center (NSIDC). These data are generated from the Special Sensor Microwave/Imager (SSM/I) and offer a different view of remotely sensed sea ice. The filenames of the SSM/I ice concentration maps generated with the NASA Team algorithm are dated. For example, if the map is of date August 14 1998, then the filename of the map is 980814.TNN. Our SSM/I database manager thus is able to locate the corresponding SSM/I file by using the ancillary data of the image. It then extracts the needed data point by geo-coordinates and retrieves the ice concentration percentage. The information has been incorporated into the knowledge base by a symbolic fact *ssmicon* and six rules that involving that particular fact.

Currently, we are working on ingesting SSM/I ice concentration maps generated with the CalVal algorithm in GRIB format by the Fleet Numerical Meteorology and Oceanography Center (FNMOC) of the U.S. Navy.

### Climatology Data (SIGC Data)

The SIGC data set is a statistical computation describing extent and coverage of sea ice in specified area of the Arctic and Antarctic Oceans. These data (53 files) were compiled from 23 years (1972-1994) of NIC Arctic and Antarctic sea ice analyses, and are important for sea ice classification as experts can draw inferences from past the climatological history of a region to help improve classification. With regards to ARKTOS, the climatology data will be used when (1) various data and knowledge sources cannot arrive at a confident classification, and (2) for specific regions of which sea ice distributions can be inferred from past climatology data—for example, when the distributions over the years can be observed in a statistically significant trend. Currently, the climatology data is not part of ARKTOS' operation due to the on-going format and medium revision phase of the data.

### HUMAN EXPERT KNOWLEDGE

The last knowledge source that ARKTOS uses is human expertise. To obtain the rules, we conducted multimedia interviews (oral, computer-based, image-based interactions) in various settings with sea ice geophysicists and photo-interpreters [1]. In addition to rules, human expert knowledge was also used in the design of attributes and determination of threshold values for the generation of symbolic facts.

The following is one of the rules used in ARKTOS:

*Rule=71; If return is dark, mottled and brighter then it might be MY; winter true, return dark, mottled true, brighter true; old\_ice; 0.7*

Each rule has five items: identification number, text description, list of antecedents, the classification, and the weight. Thus, the above rule's identification number is 71. The list of antecedents specifies that for the rule to fire, the attribute *winter* must be true, *return* must be dark, *mottled* must be true, and *brighter* must be true. When the rule fires, it will introduce a belief that the feature is old ice with a weight of 0.7.

### DEMPSTER-SHAFER BELIEF SYSTEM

After the various data and knowledge sources are ingested and with information extracted and fused into the expert system, we use a Dempster-Shafer belief system to perform the classification. Once invoked, the system accumulates the evidence systematically and handles conflicting hypotheses to finally arrive at a consistent conclusion regarding the fuzzy classification of an event, or a sea ice feature in our case. This module is essentially where all the different sources of data and knowledge are fused to give ARKTOS a consistent and intelligent reasoning capability. Interested readers are referred to Artificial Intelligence textbooks for a treatment of the Dempster-Shafer belief system.

### CONCLUSIONS

Fig. 2 shows the block diagram of the ARKTOS system. The pre-processing, segmentation, and facts extraction modules comprise the data mining part of transforming the image into feature-based, symbolic facts. Then, during the classification phase, the multisource data and knowledge bases are used to support the expert system to assign classes with varying confidence to the features. Fig. 2 also shows a GUI module that allows users such as photo-interpreters and geophysicists to analyze the results and refine the definition of facts and rules.

In conclusion, we have described the multisource data and knowledge fusion effort and modules of ARKTOS. We have discussed modules that store, select, translate and interpret data or knowledge, maintain consistency among input and output channels through the use of attributes and facts, establish inter-source information coordination using dates and geo-locations, and fuse the interpreted information using the Dempster-Shafer belief system in our rule-based expert classification.

Currently, ARKTOS is stable and undergoing fine-tuning to be operationalized at the NIC.

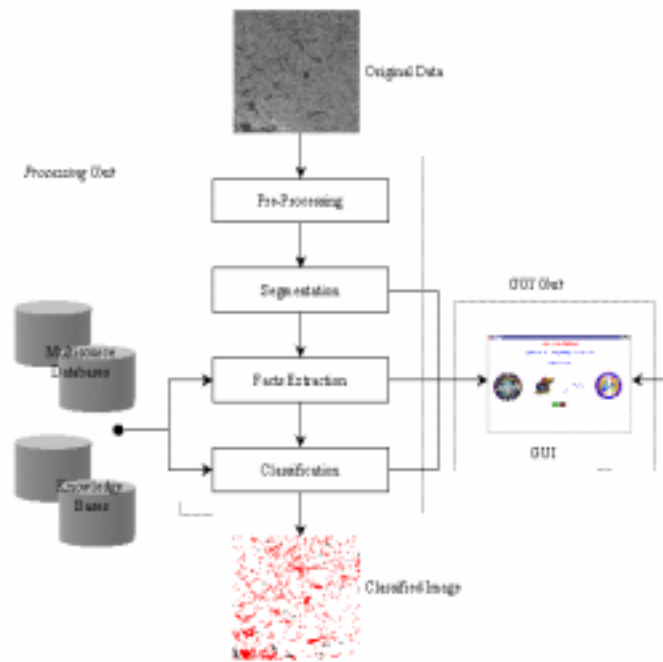


Fig. 2 The block diagram of ARKTOS.

### REFERENCES

- [1] L.-K. Soh, C. Tsatsoulis, T. Bowers, and A. Williams, "Representing Sea Ice Knowledge in Dempster-Shafer Belief System," IGARSS'98, Seattle, WA, pp. 2234-2236.