Workshop on Intelligent Techniques for Spatio-Temporal Data Analysis in Environmental Applications, Chania, Greece, 1999

INTELLIGENT FUSION OF MULTISOURCE DATA FOR SEA ICE CLASSIFICATION

Costas Tsatsoulis[1], Leen-Kiat Soh[1], Cheryl Bertoia[2], and Kim Partington[2]

[1]Information and Telecommunication Technology Center
Department of Electrical Engineering and Computer Science
The University of Kansas
Lawrence, KS 66045
USA
{tsatsoul, lksoh}@ittc.ukans.edu

[2]U.S. National Ice Center
FOB#4, 4251 Suitland Road
Washington, D.C. 20395
USA
{bertoiac, partingtonk}@natice.noaa.gov

ABSTRACT

In this paper we describe ARKTOS, a system that uses a Dempster-Shafer rule base to integrate data from multiple sources in order to classify sea ice. ARKTOS analyzes SAR imagery to generate a feature set that describes the image. Next it fuses the SAR-extracted features with digital grid climatology data and sea ice concentration data extracted from SSM/I imagery. The fusion is achieved by a set of Dempster-Shafer rules that use the multisource data to calculate belief for the various sea ice classes. The result is a fully automated system that ingests multisource data and outputs classified images of sea ice of the Beaufort Sea. ARKTOS is currently installed at the U.S. National Ice Center and at the Canadian Ice Services and is integrated in the operations flow of these organizations.

1. INTRODUCTION

Remote sensing of the polar oceans has important applications which concern meteorology and global climate. In addition, scenes of sea ice must be immediately analyzed for navigational purposes concerning oil drilling and exploration in the polar regions. The growth and decay of sea ice in the polar regions indicate changes in global weather patterns; and because the ice must be constantly monitored to recognize these changes, satellite remote sensing is the most efficient tool.

Since the launchings of the ERS and RADARSAT systems, enormous amounts of data have been made available to the scientific and operational communities and need to be processed in a timely and accurate fashion. Simple computer algorithms can produce quick classifications, but they lack the accuracy of the slower human ice classification expert. Much of the classification currently being done consists of an algorithmic processing of the data followed by a human visual check of the algorithmic results; consequently, due to speed constraints, a large part of the data is not being classified at all. The use of human experts is clearly not feasible for real-time processing of sea ice imagery; however, the knowledge used by these experts *can* be codified in rules and heuristics to assist in a quicker, automated classification.

The ARKTOS system presented here fuses SAR imagery, digital grid climatology and classified SSM/I images of the polar regions using the knowledge of human sea ice classification experts implemented as Dempster-Shafer rules to classify sea ice imagery. ARKTOS classifies sea ice into four major categories: old ice (ice thicker than 2m), first-year ice (ice of thickness between 0.3m and 2m), open water (ice thinner than 0.3m) and fast ice (ice connected to land). A diagram of the operations of ARKTOS is shown on Figure 1.

2. SEGMENTATION OF THE SAR IMAGE

ARKTOS receives RADARSAT images from many of the ground stations in the Northern hemisphere (Alaska SAR Facility in the Unites States of America, West Freugh in Scotland, Trømso in Norway, and Gatineu in Canada). Since each ground station stores images in its own unique format, we first translate the image and store it in an internal ARKTOS format which allows us to analyze images from any ground station in the same way. ARKTOS images are simply one-byte gray scale images with a land mask overlaid so that the algorithms run on sea ice only and not on any land areas. The format includes a text file that contains ancillary information about the image, such the date and time it was taken, the coordinates of the image, the satellite pass, the station that captured the data, and so on.

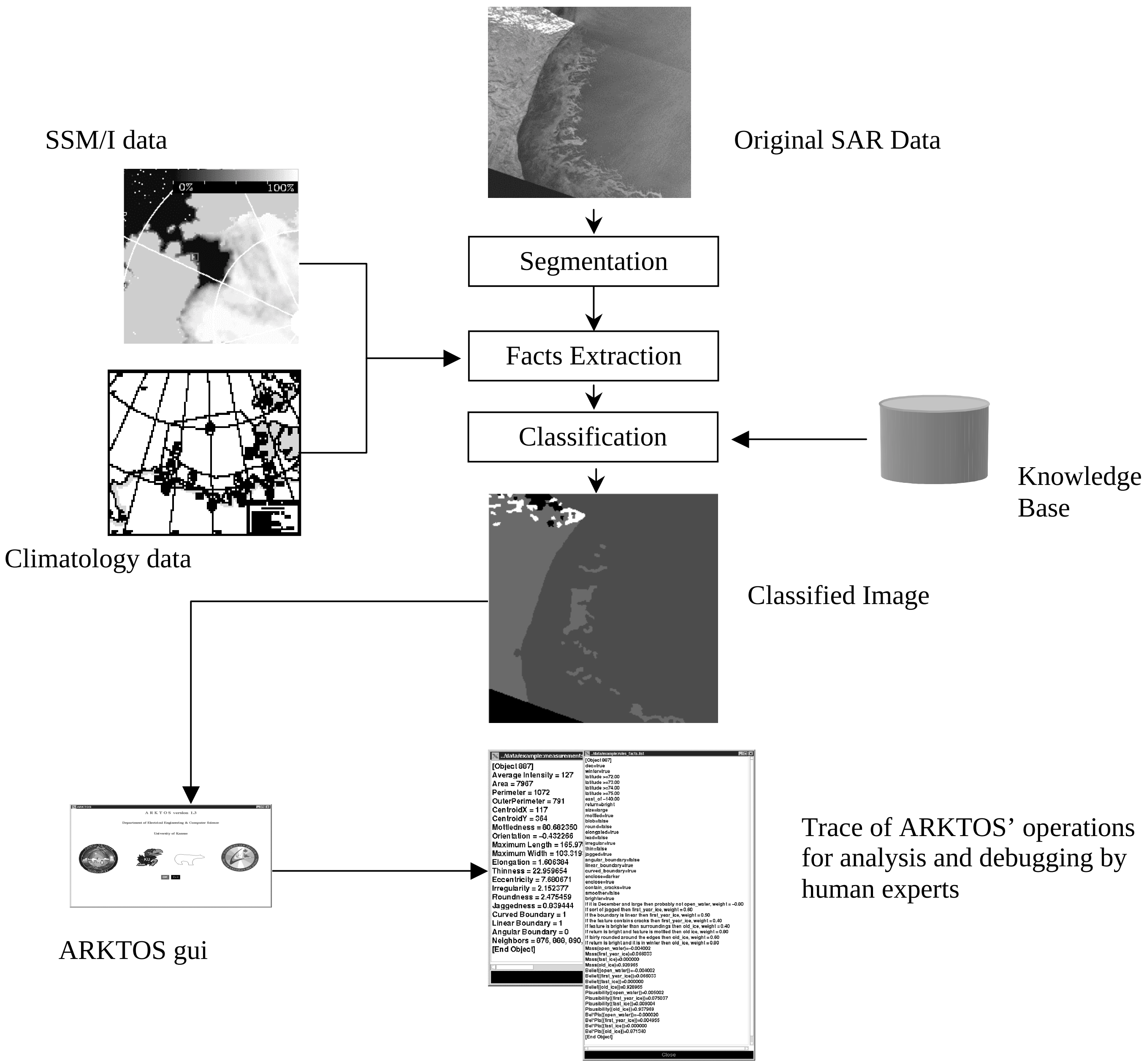

Figure 1: ARKTOS diagram of operation

Next, ARKTOS segments the image into features, i.e. areas of neighboring pixels with a common characteristic, using the Watershed algorithm (Gauch, 1999). Very briefly, the Watershed algorithm is based on the idea of considering an image as a topographic surface. The image intensity is considered as an altitude, and the watershed algorithm is applied on the gradient image where homogeneous regions have a low gradient, thus can be viewed as the "crater" of a region, and region boundaries have a high gradient, thus forming the border of the "crater." Using the local minima (i.e. the areas of low gradient) as a starting point, the algorithm merges together areas whose gradient boundary is below a certain threshold, creating what the developers of the algorithm call "catchment basins." When the process is finished, the image is separated into nonintersecting segments.

Since Watershed tends to over-partition an image, in our work we have slightly modified it by adding simple heuristics that allow some segments to be merged. For example, if the gradient of two neighboring segments is too small and their average gray value similar, we merge them. We also merge small (less than 100 pixels) segments with neighboring ones. Using these and some other heuristics allows us to segment SAR imagery into a reasonable collection of sea ice features.

3. GENERATION OF FACTS

After the image has been segmented, ARKTOS computes a set of numerical descriptors for each feature identified. These descriptors are used to later generate symbolic attributes or *facts* of the features. These facts were derived from the knowledge used by sea ice experts to classify sea ice. For example, since the shape of an ice floe is fundamental in its classification by experts, ARKTOS computes attributes such as the area of a feature, whether its sides are linear or curved, whether the floe has a jagged or smooth edge, and so on. ARKTOS currently generates and uses 17 numeric feature attributes, some of which we list and describe below:

1. *area*

Measures the number of pixels within a feature.

2. *average_intensity*

Measures the average intensity of all pixels within a feature.

3. *perimeter*

Measures the number of boundary pixels of a feature.

4. *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*.

5. *centroid*

Stores the pixel coordinates $\langle \mu_x, \mu_y \rangle$ of the center of mass of a feature.

6. *orientation*

Used to gauge how a feature aligns in the image, which is useful for computing the feature's bounding rectangle. We used the equation described in (Jain, 1989):

$$orientation = \frac{1}{2}\tan^{-1}\left[\frac{2\overline{\mu}_{1,1}}{\overline{\mu}_{2,0}-\overline{\mu}_{0,2}}\right]$$

where

$$\mu_{0,2} = \sum_{all\ y}\left(y-\mu_y\right)^2 ,$$

$$\mu_{2,0} = \sum_{all\ x}\left(x-\mu_x\right)^2 \text{ , and}$$

$$\mu_{1,1} = \sum_{all\ x}\sum_{all\ y}\left(x-\mu_x\right)\left(y-\mu_y\right) .$$

7. *roundness*

Measures how circular a feature is. First, for each *boundary* pixel, p, we compute its distance from the centroid $\langle \mu_x, \mu_y \rangle$:

$$D\left(p,\langle \mu_x,\mu_y\rangle\right) = \sqrt{\left(x_p-\mu_x\right)^2+\left(y_p-\mu_y\right)^2} .$$

Then we compute the standard deviation of all such distances and equal roundness to the standard deviation:

$$roundness = std\_dev\left[D\left(p, \left\langle \mu_x, \mu_y \right\rangle\right)\right].$$

8. *irregularity*
Measures how irregular a feature is. The approach is based on two criteria: *area_porosity* and *perimeter_porosity*:

$$area\_porosity = \frac{max\_width \bullet max\_length}{area}$$, and

$$perimeter\_porosity = \frac{\max(outer\_perimeter, perimeter)}{\min(outer\_perimeter, perimeter)}.$$

Note that *area_porosity* will be high if the feature is branchy, and if the feature has many holes and 1-pixel-thick structures, *perimeter_porosity* will be high. We compute

$$irregularity = area\_porosity \bullet perimeter\_porosity.$$

The higher the value, the more irregular the feature.

These numerical descriptors of a feature are used to generate symbolic facts, as suggested by the experts. It is these facts that are used in the Dempster-Shafer rule system for classification. Table 1 lists some of these facts and also gives their possible discrete values.

Attribute	**Possible Values**
return	`return=black, return=dark, re-turn=gray, return=bright`
size	`size=small, size=medium, size=large size=huge`
blob	`blob=true, blob=false`
round	`round=true, round=false`
elongated	`elongated=true, elongated=false`
thin	`thin=true, thin=false`
jagged	`jagged=true, jagged=false`
curved_boundary	`curved_boundary=true, curved_boundary=false`
linear_boundary	`linear_boundary=true, lin-ear_boundary=false`
angular_boundary	`angular_boundary=true, angular_boundary=false`
enclose	`enclose=true, enclose=false, en-close=darker`
smoother	`smoother=true, smoother=false`
brighter	`brighter=true, brighter=false`

Table 1: Some ARKTOS attributes with their possible discrete values

As can be seen from Table 1, some facts about a feature are directly related to the numeric attributes computed. For example, based on discussions with experts, a feature with area less than 200 pixels is considered to be of small size, a feature with area between 201 and 1600 pixels is medium, a feature between 1601 and 25000 pixels is large, and still larger features are considered huge. Other facts are based on comparison between values of two or more neighboring features. For example, “brighter” is true is the average intensity (“return”) of a feature is at least 20% higher than that of all of its neighbors.

Finally, ARKTOS generates a small set of facts that describe the date when the image was taken and the area covered by the image. These parameters are very important since SAR images of sea ice look very different during the Arctic summer and winter, and sea ice has different characteristics in different areas of the Arctic.

4. INTEGRATION OF OTHER DATA SOURCES

During our interviews of sea ice experts it quickly became obvious that classification could not be based only on SAR images, but many clues came from other data sources. Currently ARKTOS fuses two other data sources together with SAR imagery to achieve classification. The first data source is digital climatology in grid format. The climatology takes the form of two sources: historically compiled data and the last ice chart, both generated by the U.S. National Ice Center (NIC).

Historical climatology is based on ice concentration charts generated weekly by the NIC (NIC, 1996). Historically, all NIC sea ice analyses have been produced through the integration of remotely sensed and *in-situ* oceanographic and meteorological data. Today, sea ice analysis at the NIC is done almost exclusively with remotely sensed data. The satellites and sensors used to produce global sea ice analyses included: TIROS visible/infrared GAC/LAC/HRPT (VHRR and AVHRR) data; NIMBUS passive microwave (ESMR and SMMR) data; DMSP visible/infrared (OLS) smooth and fine data; GEOSAT altimetry data; DMSP passive microwave (SSM/I) data; ERS-1 Synthetic Aperture Radar (SAR) data; and RADARSAT SAR data. The percent utilization of each data type varies both temporally and spatially in the weekly analysis files. For example, summer ice analyses during the period 1972-1983 were based on 60% visible/infrared data, 30% microwave data and 10% conventional (aerial or *in-situ*) observations. The NIC ice charts of the previous week are also fused with the SAR-generated facts to provide a more recent historical perspective for the algorithm.

Finally, ARKTOS takes the results of sea ice classification algorithms of SSM/I passive microwave data (using either the CAL/VAL or the NASA Team algorithms) and generates facts that allow it to fuse a passive data source with SAR and historical climatology data.

After the SAR image analysis and fact extraction phases ARKTOS has available to it a set of features associated with a number of attributes (descriptors) generated from three distinct data sources.

5. KNOWLEDGE-BASED DATA FUSION AND CLASSIFICATION

We conducted three week-long interviews of sea ice experts to collect our classification knowledge and all the attributes used in classifying and describing sea ice. Our knowledge acquisition was based on three methods: protocol, blind test, and reference (Hart, 1992; Soh et al., 1998). Protocol is based on the "talk-trough" method where the expert thinks aloud while classifying the image. The experts interviewed were seated in front of a computer monitor and were asked to classify verbally an electronically displayed SAR image of sea ice. This method provided an unstructured interview environment since no questions were pre-determined. It also closely simulated the normal work environment of sea ice experts. The blind test method is based on the concept of "twenty questions," where the expert asks questions about an unseen case, thus having to verbally describe the classification process by classifying an image without seeing it. Finally, our reference interview sessions used pre-determined questions to elicit specific information. These sessions helped resolve conflicts, verify assertions, and refine the certainty and accuracy of the acquired knowledge.

All classification rules we collected were uncertain, that is, the experts always expressed a degree of belief in the classification based on certain observables. We decided to combine the uncertainty values of the classification rules using Dempster-Shafer theory. Our decision was based on several reasons: first, Dempster-Shafer seems to us to better implement human decision making but in a formal and mathematically sound framework; second, Dempster-Shafer is especially suited to decision making where there are conflicting hypotheses being supported sometimes by the same evidence; and, third, Dempster-Shafer can handle sets of classification hypotheses.

The rules are stored in an ARKTOS-specific format that allows the users of the system to edit them and fine-tune them. The human experts and operators of ARKTOS can use a graphical user interface (gui) to edit rules, change parameters, and view a trace of the system's operations for debugging and for improvement of the rule base. An example of a rule is the following:

```
rule=1; If real smooth and bright then open water;  surface=smooth, re-
turn=bright; open_water; 0.6
```

The rule is preceded by a unique number, followed by a text description of the rule, next the list of predicates with their values, the classification, and finally the mass assigned by the expert to the classification.

Other rules integrate features from different sources to generate a classification. For example:

```
rule=19; If real smooth and the SSM/I concentration is less than 30%
then open water;  surface=smooth, ssmicon<30; open_water; 0.8
```

This rule looks at the texture of a feature and, if it is smooth, looks at the SSM/I-derived ice concentration in the area. If the concentration is low, then ARKTOS can assume the feature in question is probably open water.

6. CONCLUSIONS

ARKTOS is an automated, intelligent system that receives SAR satellite data, analyzes it, generates a set of features and descriptors for each feature, integrates these descriptors with information extracted from classified SSM/I data and digital climatology maps, and finally uses expert rules and Dempster-Shafer theory to classify the sea ice in the image.

As far as we know ARKTOS is the only fully automated system that classifies sea ice in an operational environment. ARKTOS has been installed and is being tested and used at the U.S. National Ice Center and at the Canadian Ice Services. Both organizations are using ARKTOS to classify sea ice in the Beaufort Sea.

7. ACKNOWLEDGEMENTS

This work was supported in part by a Naval Research Laboratory contract N00014-95-C-6038 and in part by the Research Development Fund of the University of Kansas.

8. REFERENCES

Gauch, J.M. 1999. "Image Segmentation and Analysis via Multiscale Gradient Watersheds," *IEEE Trans. on Image Processing,* Vol. 8, No. 1.

Hart, A. 1992. *Knowledge Acquisition for Expert Systems,* second edition, New York: McGraw-Hill.

Jain, A.K. 1989. *Fundamentals of Digital Image Processing,* Englewoods, CA: Prentice Hall, 392-393.

NIC. 1996. National Ice Center, Fleet Numerical Meteorology and Oceanography Detachment, and National Climatic Data Center, *Arctic and Antarctic Sea Ice Data 1972-1994*, CD-ROM Version 1.0, July.

Soh, L.-K., C. Tsatsoulis, T. Bowers, and A. Williams. 1998. “Representing Sea Ice Knowledge in a Dempster-Shafer Belief System,” *Int. Geoscience and Remote Sensing Symp.*, 2234-2236.