Oil Spill Detection in Radarsat and Envisat SAR Images

Anne H. S. Solberg, *Member, IEEE*, Camilla Brekke, and Per Ove Husøy

Abstract—We present algorithms for automatic detection of oil spills in synthetic aperture radar (SAR) images. The algorithms consist of three main parts, namely: 1) detection of dark spots; 2) feature extraction from the dark spot candidates; and 3) classification of dark spots as oil spills or look-alikes. The algorithms have been trained on a large number of Radarsat and Envisat Advanced Synthetic Aperture Radar (ASAR) images. The performance of the algorithm is compared to manual and semiautomatic approaches in a benchmark study using 59 Radarsat and Envisat images. The algorithms can be considered to be a good alternative to manual inspection when large ocean areas are to be inspected.

Index Terms—Classification, feature extraction, oil spill detection, synthetic aperture radar (SAR).

I. INTRODUCTION

MARINE pollution arising from illegal oily discharges from tank cleaning or bilge pumping represents a serious threat to the marine environment. Discharges from ballast water or tank cleaning concern mainly tankers, whereas discharges from engine room effluent concern all types of ships. Due to such operations, large amounts of oil are deliberately pumped into the marine environment every day. Oil spills correlate well with the major shipping routes; they also often appear in connection with offshore installations.

The combined use of satellite-based synthetic aperture radar (SAR) images and aircraft surveillance flights is a cost-effective way to monitor large areas and catch the polluters. Radarsat and Envisat Advanced Synthetic Aperture Radar (ASAR) images enable covering large areas; however, aircraft observations are needed to prosecute the polluter and, in certain cases, to verify the oil spill.

Oil spills appear as dark areas in the SAR images because the oil dampens the capillary waves of the sea surface. A major part of the oil spill detection problem is to distinguish oil spills from other natural phenomena (look-alikes) that dampen the short waves and create dark patches on the surface. An oil spill may result in several oil slicks. By the term "oil slick," we will denote a region on the sea surface covered by oil. We distinguish between the terms "oil spill" and "oil slick" to be able to describe each dark region in the image that the oil spill consists of.

A review of algorithms for automatic detection of oil spills can be found in [1]. Several of the papers [2]–[4] describe a methodology consisting of dark spot detection followed by feature extraction and classification. Solberg *et al.* [4] apply an adaptive thresholding algorithm for dark spot segmentation. Kanaa *et al.* [5] use hysteresis thresholding. An edge detection approach based on the Laplace of Gaussians or Difference of Gaussians is presented in [6] and [7]. The use of wavelets for ocean feature detection is described in [8] and [9]. QinetiQ's dark spot algorithm uses a constant false alarm rate algorithm followed by clustering and Hough transform to identify linear targets [10]. Barni *et al.* [11] propose an algorithm based on fuzzy clustering. A method using mathematical morphology for oil spill segmentation is presented by Gasull *et al.* [12]. Even though a variety of methods are applied, the common approach is to detect all suspicious slicks and to preserve slick shapes.

In [1], a discussion of previous approaches to oil spill feature extraction is included. Most of the features used for slick feature extraction in the literature are typically covered by the following types: geometry and shape of the segmented region; physical characteristics of the backscatter level of the spot and its surroundings; contextual features describing the slick in relation to its surroundings; and textural features.

Various classifiers have been applied to classify a slick as oil or look-alike. Fiscella *et al.* [2] applied a Mahalanobis classifier. Solberg *et al.* [4] combined a statistical classifier using a Gaussian model and several subclasses with a rule-based modification of prior probabilities to take into account feature combinations that are indications of certain scene conditions. They tested their method on 11 oil spills and six look-alikes. A neural-network approach is described by Del Frate *et al.* [3]. Their data set consisted of 71 oil slicks and 68 look-alikes, and they used leave-on-out for error estimation. A recent study based on fuzzy logic is given by Keramitsoglou *et al.* [13], who used nine images for training and 26 images for testing. A comparison of these different oil spill detection approaches in terms of classification accuracy is difficult. They use different data sets and different classifiers, so the reported classification accuracy cannot be directly compared. For an automatic algorithm that should operate under all wind conditions, including low wind where a high number of look-alikes can be expected, the false alarm ratio becomes very important. By using, for example, loss functions, the number of oil spills classified as look-alikes can be lowered, but at the cost of a higher false alarm ratio. In this paper, we test the performance of the algorithm on a test set of several thousands of look-alikes.

Comparing the results of the automatic algorithm to manual inspection by a trained operator is very interesting. A significant contribution of this paper is a benchmark study involving 59 Radarsat and Envisat SAR images that compares the algorithm to the manual detection done as part of the operational oil spill detection service at the Kongsberg Satellite Station, Tromsø, Norway, and aircraft verification.

In this paper, we present a methodology for automatic oil spill detection in Radarsat and Envisat SAR images. The algorithms are trained on 70–80 images from each sensor, and they are benchmarked against aircraft verifications and manual detections.

II. SAR IMAGING OF OIL SPILLS

A SAR sensor is the most efficient satellite sensor for oil spill detection, though it does not have capabilities for oil spill thickness estimation and oil type recognition. A discussion of other sensors used for oil spill imaging is given in [1]. Oil spills dampen the Bragg waves on the ocean surface and reduce the radar backscatter coefficient, thus creating dark spots in the SAR image. Oil spill look-alikes are phenomena that result in dark areas in the SAR image. Oil spills may include all oil-related surface films caused by oil spills from oil rigs, leaking pipelines, passing vessels as well as bottom see pages, whereas look-alikes include natural films/slicks, algae, grease ice, threshold wind speed areas, wind sheltering by land, rain cells, shear zones, internal waves, etc. [14]. A trained human interpreter is able to discriminate between oil spills and look-alikes based on experience and prior information concerning location, differences in shape and contrast, and weather information. Knowledge about ship lanes and oil rig positions can be extracted from a database, but ship and oil rig positions can also be derived from the SAR image (ships of a certain minimum size). Information about the wind level is very important to consider for oil spill classification. The wind level influences the backscatter level and the visibility of oil spills and look-alikes. The visibility of slicks in SAR images is discussed in [15] and [16]. In wind speeds less than 3–5 m/s, the probability of observing look-alikes is high. Wind speeds from 5 to 10 m/s produce fewer dark spots from local low-wind areas. As the wind increases, the contrast between the oil and the surrounding sea will decrease. In high wind, only thick oil will be visible, and look-alikes are rare. The upper limit for observing oil in the SAR image is not known exactly. In an operational oil spill detection service at the Kongsberg Satellite Station, an upper limit of 15 m/s is used.

III. OIL SPILL DETECTION APPROACH

An overview of the oil spill detection approach is given in Fig. 1. Preprocessing consists of land-masking the images by converting a land mask to the SAR image grid to avoid resampling the speckle pattern, followed by range-dependent scaling and conversion to radar brightness values (beta naught values). The land mask is obtained from a shoreline database [17].

The radar backscatter from the ocean depends strongly on the incidence angle, yielding a much stronger signal at low incidence angles. The incidence dependency depends on the
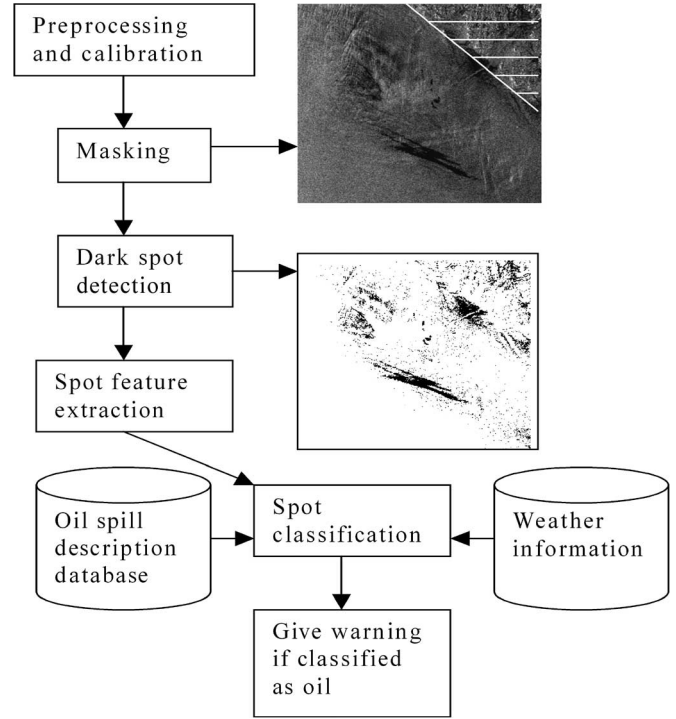

Fig. 1. Overview of the oil spill detection approach.

wavelength, polarization, and wind and weather conditions on the ocean surface. Instead of trying to estimate and compensate for the weather-dependent effects, we have opted for the simple solution of using a standard incidence angle compensation for all weather conditions. Two different correction functions were identified in a small experimental study: one for Radarsat and one for ASAR. For ASAR, a $\tan^4(\alpha)$ term is used, where α is the incidence angle; whereas for Radarsat, a $\tan^2(\alpha)$ is sufficient to yield a visually satisfactory range intensity for the relevant scenes.

A. Dark Spot Detection and Segmentation

The first step in the oil detection algorithm is to segment out oil spill candidates. We have chosen to segment out all dark regions in the image and later discriminate between oil spills and look-alikes based on features computed for each region. This approach can result in a large number of dark areas being segmented, particularly for low-wind conditions where look-alikes are frequently observed.

Sensor-specific modules for dark spot detection have been developed for Radarsat ScanSAR and Envisat Wide Swath (WS) ASAR images. Sensor-specific modules were necessary because of different spatial resolution and contrast. Radarsat ScanSAR Narrow full-resolution images (HH polarization) have a resolution of approximately 50 m and a pixel spacing of 25 m. Envisat WS images (VV polarization) have a resolution of approximately 150 m and a pixel spacing of 75 m.

Initial segmentation experiments with the algorithm developed for the European Remote Sensing (ERS) images [4] showed that a single-scale approach did not perform well in segmenting both large and small regions. Thus, a multiscale approach was developed.

TABLE I
PARAMETER VALUES FOR THRESHOLDING

PMR intervals	Homogeneity category (k)	Thresholds ΔdB_k in dB (Envisat)	Thresholds in dB (Radarsat)
PMR > 0.15	1	4.0	4.0
PMR $\in [0.05, 0.15]$	2	2.7	3.0
PMR $\in [0.04, 0.05]$	3	2.4	2.0
PMR $\in [0.03, 0.04]$	4	1.5	1.8
PMR $\in [0.015, 0.03]$	5	1.3	1.5
PMR < 0.015	6	1.0	1.0

First, an image pyramid is created by averaging pixels in the original image. From the original image, the next level in the pyramid is created with half the pixel size of the original image, and so on. Adaptive thresholding is then applied to each level in the pyramid.

The following adaptive thresholding algorithm is used to segment each level in the pyramid.

1) For each pixel i, compute the mean value μ and the power-to-mean (PMR) value σ/μ in a local window of size W, where σ is the standard deviation.
2) Compute the *homogeneity category* k from PMR as described in Table I.
3) Get threshold value ΔdB_k (in decibels) given the *homogeneity category* in Table I.
4) Set $T_i = \mu - \Delta dB_k$.
5) Threshold pixel i with the computed value of T_i.

The threshold is thus set adaptively based on estimates of the roughness of the surrounding sea. The motivation for this is given as follows. In low wind with many look-alikes, the PMR will be high, and a high contrast between slicks and their surroundings can be expected. As the wind increases, the PMR value will decrease, and the expected contrast between the oil and the surrounding sea will also decrease.

The parameter values in Table I were found by experiments on the training data set (see Section III-C).

It is possible to replace the homogeneity category with wind predictions or estimated wind levels. For the data set that we worked with, no wind predictions were available, and we did not have access to an algorithm for wind speed estimation based on the SAR image.

If μ and the PMR value are recomputed for every single pixel, then this procedure will be fairly slow for a 400×400 km^2 image (Envisat WS). The segmentation step is the most computationally intensive step of the algorithm. To speed up the processing, two skip factors are introduced. Skip factor s_1 is used for thresholding (the adaptive window is moved s_1 pixels for each threshold computation), whereas skip factor s_2 is used when computing the mean and PMR value inside the window (only every s' pixel is included in the mean and variance computations). The following parameter values are used: $W = 121$; $s_1 = 5$; and $s_2 = 11$.

After segmenting each level in the pyramid, the different levels are merged using a very simple procedure. All segmented pixels in the highest resolution level are kept, and the lower resolution levels are used to fill holes in the segmentation result using a logical operator. The Radarsat segmentation module uses three pyramid levels, whereas the Envisat module uses two levels.
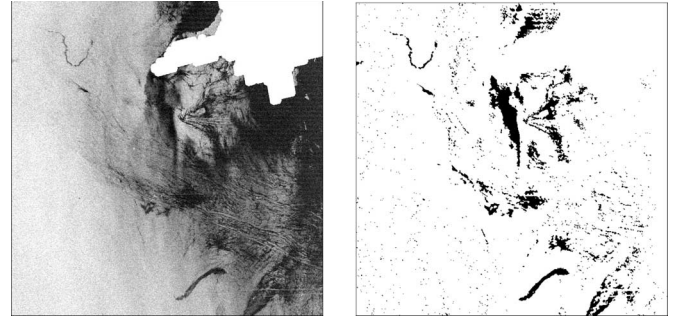

Fig. 2. Part of the Envisat image from July 21, 2003, and the corresponding segmentation result (copyright ESA/KSAT/NR).

The segmented image will not only contain all oil spills, as well as a high number of look-alikes. The segmentation step works well in general; however, in some cases, thin linear slicks can be fragmented. Fig. 2 shows an example of an image and the corresponding segmentation result.

B. Dark Spot Feature Extraction

After segmentation, "region objects" are formed for all objects in the segmented image. Small objects (< 20 pixels) are processed no further. The subsequent processing considers only region objects. For each region, a set of features is computed. These features are later used to classify each slick as oil or look-alike.

The feature set consists of a mix of features specially developed for oil spill detection and traditional descriptors from the image analysis literature.

The features can be grouped into descriptors for *shape*, *contrast*, *slick surroundings*, and *slick homogeneity*. We previously performed a study regarding oil spill features in the ERS images [4], and the features used here are mainly from [4], but some of them have been modified.

The following features are used.

1) Shape features:
   a) *Slick complexity* is defined by $C = P^2/A$, where P is the perimeter, and A is area of the region.
   b) *Slick width* is the ratio between the area of the region and the width of the branches of the skeleton of the region.
   c) *Slick area* is the size (in number of pixels) of the region.
   d) *Slick moment* is the first invariant planar moment [18], which is defined as $\phi_1 = \eta_{20} + \eta_{02}$, where η_{pq} are normalized central moments, and $\eta_{pq} = \mu_{pq}/\mu_{00}$ and μ_{pq} are the nonnormalized central moments.
2) Contrast features:
   a) *Slick local contrast* is the difference between the mean value of the slick and the mean value of a larger window surrounding the slick.
   b) *Border gradient* is the mean of the magnitude of gradient values of the region border area. The Sobel operator is used to compute the gradient.
   c) *Smoothness contrast* $(N_r/G_r)/(N_b/G_b)$ is defined as the ratio between the number of region pixels N_r and the sum of the region pixel gradient values G_r and the

ratio of the number of background pixels N_b and the sum of the background gradient values G_b.

3) Homogeneity features:

  a) *Surrounding PMR* is computed as the PMR ratio σ/μ in a large window containing slick surroundings.

  b) *Slick PMR* is computed as the PMR ratio of the slick.

4) Slick surroundings:

  a) *Number of detected spots in the scene.*

  b) *Number of neighboring spots in a large window centered at the region.*

  c) *Number of neighboring spots in a small window centered at the region.*

  d) *Distance to ship/rig.* A simple point detector to detect ships or oil rigs has been implemented. To detect point targets, pixel cliques (a clique consists of two or more neighboring pixels) within a center window of size 3×3 is compared to the mean value of a larger window of size 15×15. If the difference between the clique with the largest mean value and the mean of the large window is greater than 7 dB, the center location is said to be a point target. In every region of spot detection, the distance to the closest point target is computed.

C. Dark Spot Classification

In designing a classifier to discriminate between oil spills and look-alikes, several important factors are considered. One factor is the expected number of oil spills compared to the expected number of look-alikes. The number of look-alikes depends on the wind speed and location. With the given segmentation approach, we sometimes observe several hundreds of look-alikes in a single scene during low-wind conditions. In high wind, only a few look-alikes are expected. The expected number of oil spills does not depend strongly on wind speed. In low wind, even very thin oil films are visible in the SAR image, whereas in high wind, only thicker oil films are visible. The expected number of oil spills in a scene might also depend on the location. Oil spills are commonly seen near oil rigs or major shipping lanes. Fig. 3 shows the number of reported oil spills in the Baltic Sea during 1990–2002. Note that this information is not normalized in terms of monitoring frequency, so it cannot be directly used to calculate the yearly expected number of oil spills at a given location. Studying the spatial distribution and temporal trends for oil spills is beyond the scope of this paper. For further information, see [19].

The difference between the expected number of oil spills and the number of look-alikes will influence the estimates of the corresponding class-conditional probability distributions. A training set consisting of, for example, 100 images is likely to contain thousands of look-alikes and about 100 oil spills. We need to consider this when estimating the variance of the distributions.

1) Basic Classifier: In this section, we describe our basic classifier for oil spill classification. It is given in [4], but it is included here for easier reference. For each of the detected dark spots, a number of features are computed and collected in the feature vector $\mathbf{x}_i$. These features are constructed such that they
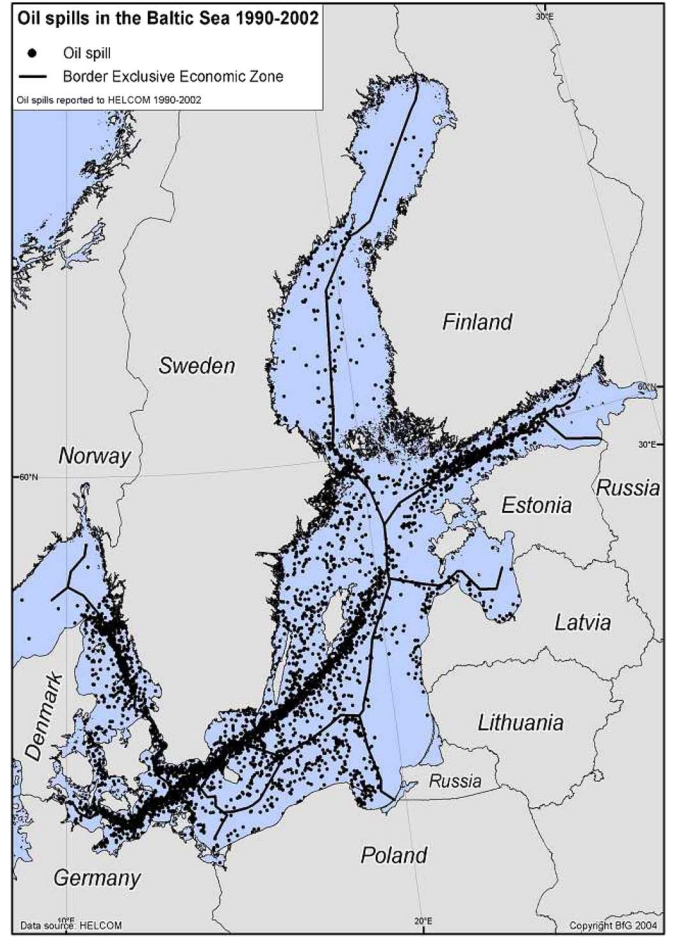

Fig. 3. Oil spills in the Baltic Sea during 1990–2002 reported to the Helsinki Commission (copyright BfG 2004).

typically will be different, depending on whether the dark spot is an oil spill or a look-alike. However, the behavior of these features will change with different wind levels. Denote the wind level by w.

A prior distribution and a probability density for the features are combined through Bayes' theorem to obtain the *posterior* probability for a detected spot being an oil spill. Let c be the unknown class membership of a detected spot. Then

$$
\begin{aligned}
\Pr(c = o | \mathbf{x}_i, w) &= \frac{\pi_o(w) f_{o,w}(\mathbf{x}_i)}{\pi_o(w) f_{o,w}(\mathbf{x}_i) + (1 - \pi_o(w)) f_{l,w}(\mathbf{x}_i)} \\
&= \frac{\pi_o(w) \frac{f_{o,w}(\mathbf{x}_i)}{f_{l,w}(\mathbf{x}_i)}}{\pi_o(w) \frac{f_{o,w}(\mathbf{x}_i)}{f_{l,w}(\mathbf{x}_i)} + (1 - \pi_o(w))}.
\end{aligned} \tag{1}
$$

$\pi_o(w)$ and $\pi_l(w)$ are the prior models for the probability that a detected spot is oil or look-alike given wind level w. $f_{o,w}(\mathbf{x}_i)$ and $f_{l,w}(\mathbf{x}_i)$ are the probability densities for the observed features $\mathbf{x}_i$ in classes $o =$ oil spills and $l =$ look-alikes, respectively.

In the following sections, we describe how the class-conditional probability densities are computed, the use of prior models for the expected number of oil spills and look-alikes in an image, the use of loss functions to model false positives and false alarms, and the use of rule-based corrections of the prior probabilities.
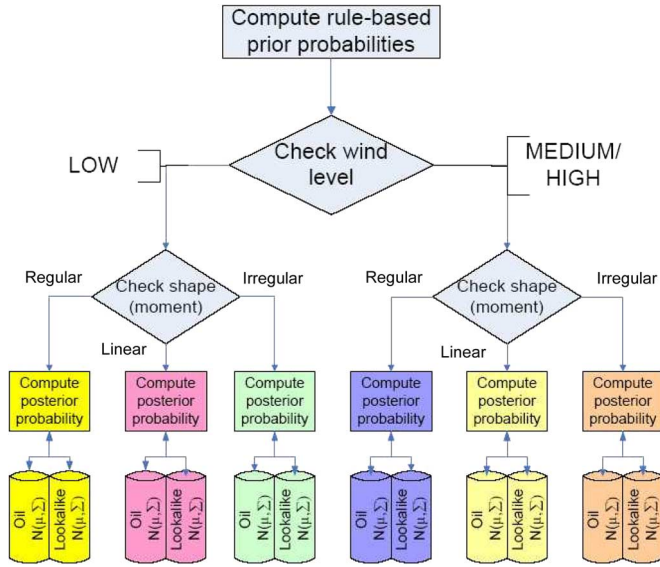
Fig. 4. Subclass structure based on wind and shape.

TABLE II
EXPECTED NUMBER OF LOOK-ALIKES IS GIVEN BY $1/(1 - \alpha_w)$, WHERE α_w IS THE CORRESPONDING PARAMETER IN THE GEOMETRIC DISTRIBUTION FOR THE NUMBER OF LOOK-ALIKES

Wind level	Expected no. look-alikes
Low	10000
Low-to-moderate	100
Moderate	20
High	7

probability densities. These features were selected in previous studies [4].

b) Prior model for the number of slicks: Prior distributions for the presence of oil spills can, in its simplest form, be specified through the number of oil slicks in the scene. In particular, we will assume that the probability of k oil slicks in the scene is given by

$$\pi(k) = \alpha^k (1 - \alpha) \qquad (2)$$

where $1/\alpha$ is a parameter describing the expected number of oil slicks in the scene.

The number of detected spots is denoted by M. We assume that all the present oil slicks will be detected as dark spots. Given the number of oil slicks M_o, the stochastic variable is $M_l = M - M_o$, which is the number of look-alikes that will appear in the scene. We will assume that

$$\Pr(M_l = m | M_o = k, w) = \alpha_w^m (1 - \alpha_w)$$

is a geometric distribution with the parameter, depending on the wind level given by the dependence on w. In particular, we have assumed that the wind level is divided into four categories. These categories are given in Table II, together with their corresponding values of $1/(1 - \alpha_w)$. We use the following equation as the prior $\pi_o(w)$ for an object being an oil slick:

$$\Pr(c = o | M, w) = \frac{\frac{\alpha}{\alpha_w}}{\frac{\alpha}{\alpha_w} + 1}.$$

a) Specification of class-conditional probability densities: The simplest choice of densities is to use multivariate Gaussian ones. Even within each wind level, both the oil spills and the look-alikes may vary quite a lot in shape and other features. Therefore, describing the feature density by a unimodal density, such as the Gaussian, is not appropriate. Instead, we have assumed different densities, depending on the value of a shape descriptor s (we have used the feature *slick moment*). The sample space of s was divided into four subgroups, ranging from wide slicks with regular shape to wide slicks with a complex boundary to thin linear spots. The density for class c and wind level w is then given by $f_{c,w}(\mathbf{x}_i) = f_{c,w,g}(\mathbf{x}_i)$ if s is in subgroup g. This hierarchical structure is illustrated in Fig. 4. Finally, the densities within each subgroup, which are shown as follows, are assumed to be Gaussian:

$$f_{c,w,g}(\mathbf{x}_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_{w,g}|^{\frac{1}{2}}}$$
$$\times \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \mu_{c,w,g})^T \Sigma_{w,g}^{-1} (\mathbf{x}_i - \mu_{c,w,g}) \right\}$$

where $c \in \{o, l\}$, d is the number of features, $\mu_{c,w,g}$ is the mean vector under wind level w and shape descriptor in subgroup g for class c, and $\Sigma_{w,g}$ is a diagonal covariance matrix, which is common for both classes given (w, g). A diagonal covariance matrix is used because the number of oil spills in each subclass in the training data set is small, which is typically 10–20. Equal variance for a given feature for a given oil spill subclass and its corresponding look-alike subclass is used because the look-alike classes have many times more observations than the oil spill classes, and experiments showed that using class-conditional covariance estimates resulted in classifying all slicks as look-alikes.

Only the features *slick complexity*, *slick local contrast*, *border gradient*, *smoothness contrast*, *slick PMR*, *slick width*, and *number of neighboring spots* are used to compute the

c) Loss functions: We want the classifier to find all spots with a certain probability of being oil. All spots classified as oil are then inspected by an operator. Misclassifying oil as look-alikes is considered more serious than misclassifying look-alikes as oil. This can be modeled in terms of a simple loss function. Let l_1 be the loss associated with misclassifying a true oil slick as a look-alike, and let l_2 be the loss associated with misclassifying a true look-alike as an oil slick. Then, the optimal classification for spot i is given by

$$\hat{c} = \begin{cases} o, & \text{if } \Pr(c = o | \mathbf{x}_{i,w}) > \frac{l_2}{l_1 + l_2} \\ l, & \text{otherwise.} \end{cases}$$

2) Extended Classifier: Rule-Based Corrections: The classification performance using the basic classifier was not satisfactory, mainly because a high number of look-alikes were classified as oil. A set of rules adjusting the prior probabilities based on a certain combination of features was developed.

The basic classifier resulted in a high false alarm ratio, which is unacceptable because a scene in low-wind conditions can contain hundreds of look-alikes. To overcome this, we

TABLE III
EXAMPLES OF RULES USED TO COMPUTE THE RULE-BASED MODIFICATIONS OF THE PRIOR PROBABILITIES. THE FEATURE ABBREVIATIONS ARE: DIST = DISTANCE TO SHIP/RIG, NLN = NUMBER OF NEIGHBORING SPOTS IN A LARGE WINDOW, AREA = SLICK AREA, COMPLEXITY = SLICK COMPLEXITY, PMRSURR = SURROUNDING PMR, LCONT = SLICK LOCAL CONTRAST, NSN = NUMBER OF NEIGHBORING SPOTS IN A SMALL WINDOW, MOM = SLICK MOMENT

Increase prior probability for oil spills if
$DIST < 20$ AND $NLN = 0$ AND $AREA > 20$
$DIST < 10$ AND $NLN = 0$ AND $COMPLEXITY < 20$
Increase prior probability for look-alikes if
$AREA < 20$ AND $PMRSURR > 0.05$
$NSN > 5$ AND $PMRSURR > 0.05$ AND $LCONT < 3.0$
$WIND = LOW$ AND $NLN > 5$ AND $MOM < 1$

TABLE IV
CLASSIFICATION ACCURACIES ON THE ENVISAT TEST SET WITH AND WITHOUT THE RULE-BASED CONFIDENCE LEVELS

Basic classifier, no rules		
	Classified as oil	Classified as look-alike
Marked as oil	33 (89.2)%	4 (10.8%)
Marked as look-alike	3504 (28.9%)	8613 (71.1%)
Advanced classifier with rules		
	Classified as oil	Classified as look-alike
Marked as oil	29 (78.4%)	8 (21.6%)
Marked as look-alike	77 (0.7%)	12033 (99.4%)

inspected the misclassified slicks in the training data set to understand why they were misclassified. What we found was that a human operator, in many cases, could identify many such slicks as look-alikes based on a combination of features. We then studied the feature values for misclassified slicks in relation to why an expert would label them look-alike. Then, a rule for these circumstances was established. The rule specifies a multiplicative factor adjusting the Gaussian densities (or actually the ratios between the densities of oil slick and look-alikes). Establishing these rules was a tedious process for the first images; but after performing it for 10–20 images, the procedure tended to converge, and the number of false alarms was greatly reduced.

The procedure resulted in 37 rules for Radarsat imagery. Examples of the rules used are given in Table III. Three of these were used to set prior probabilities for regions close to point sources (using the DIST feature). Most of the rules (28) were used to reduce false alarms, and eight rules were used for false negatives. For Envisat, the rule set was extended with 13 rules tailored to Envisat statistics. Generally, we observed more look-alikes in Envisat images than Radarsat images (as could be expected due to the higher sensitivity of Envisat's VV polarization), making additional rules necessary.

IV. EXPERIMENTAL RESULTS

The oil spill detection system is trained and tested on Envisat ASAR and Radarsat SAR images. From 2003 to 2004, a set of 56 Envisat WS ASAR images and 71 Radarsat SAR images from European waters (mainly the Baltic and the North Sea) was used to train the algorithm. All the SAR images were processed by Kongsberg Satellite Services (KSAT).

For some of the images, the reports made by KSAT as part of their oil spill detection service were available. To label all of the images in the training set, we created a training mask for each SAR image. KSAT reports were used when available. We scanned through all images, labeling all possible oil spill candidates as either *oil spill* or *doubt*. Doubt cases were used because, in many cases, a trained operator cannot discriminate perfectly between oil slicks and look-alikes based on the SAR image alone. All slicks not marked as oil or doubt were used to train the *look-alike* class; thus, the slicks marked *doubt* were left out of the look-alike class.

We first studied the performance of the oil spill detection algorithm compared to the test masks made by manual inspec-

tion. The Envisat test set of 27 images contained 37 slicks that were manually labeled as oil spills, in addition to 12 110 look-alikes. This set of images will also be used for benchmarking against other approaches discussed in Section IV-A. Table IV summarizes the classification accuracy for the Envisat test data set. To study the effect of the rule-based adjustments of the prior probabilities, the performance is reported with and without the rules. On one hand, without the rule-based corrections, the average classification accuracy is 89% for oil spills and 71% for look-alikes, and since the total number of look-alikes is 12 110, there are 3504 false alarms. This is not acceptable. On the other hand, with the rule-based corrections, the average classification accuracy is 78% for oil spills and 99.4% for look-alikes. This results in an acceptable false alarm ratio.

A. Benchmarking Oil Spill Detection Systems

As part of the European Commission project Oceanides, a joint satellite–airborne campaign was performed during 2003 to establish a data set consisting of SAR images with associated aircraft verifications by the German and Finnish pollution control authorities. The campaign covered the Finnish and German sectors of the Baltic Sea, in addition to the German sector of the North Sea. A total of 59 Radarsat and Envisat images were acquired between July and December 2003. This campaign was organized in such a manner that KSAT downloaded the satellite images, analyzed them in near-real time, and reported possible oil spills to the Finnish and German pollution control authorities, which would check the reported locations and determine if the reported slicks were oil spills or look-alikes.

The Radarsat benchmark data set consisted of 32 Radarsat images, and the Envisat benchmark data set consisted of 27 Envisat images. The details of this benchmarking can be found in [10]. The images were analyzed by KSAT, by a semiautomatic algorithm developed at QinetiQ, and by our automatic algorithm. For the benchmark comparison KSAT, let another operator analyze the images without knowing the results of the previously reported oil slicks. The results from this operator are termed KSAT-B and are used in the benchmark tables.

The benchmarking was done without any of the persons/algorithms knowing the aircraft detections, so that none of the approaches could be tuned to the aircraft results. The joint aircraft–satellite campaign was organized in such a manner that KSAT downloaded the satellite images, analyzed them in near-real time, and reported possible oil spills to the Finnish and German pollution control authorities, which would check the reported locations. Aircraft verification of all slicks was not possible, as many of the slicks were outside the Finnish

TABLE V
PERFORMANCE IN DETECTING VERIFIED OIL SLICKS

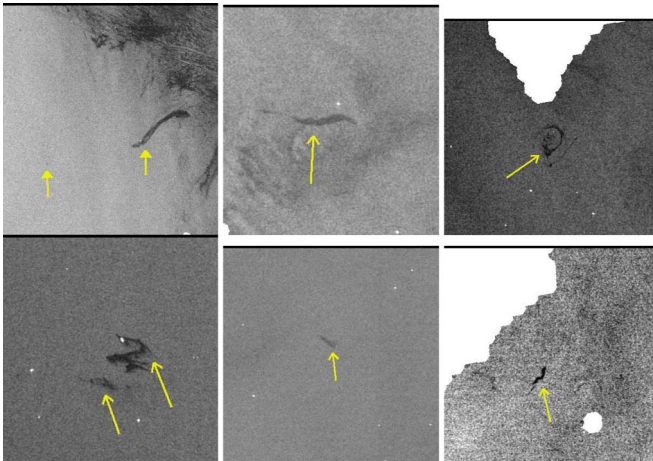|	Envisat data set	Radarsat data set
Slicks verified by aircraft	11	18
Detected by operator KSAT-B	8 (72%)	15 (83%)
Detected by algorithm	8 (72%)	14 (77%)
Detected by QinetiQ	5 (45%)	12 (66%)

Fig. 5. Envisat images with examples of verified oil slicks correctly classified as oil (copyright ESA/KSAT/NR).

or German territories, which means that the aircraft was not allowed to fly to the location, or because flight plans or satellite acquisitions sometimes had to be changed.

The performance in detecting verified oil slicks is given in Table V. The Radarsat data set contained 18 oil slicks verified by the aircraft. The KSAT-B operator found 15 of these slicks, the algorithm described in this paper found 14, and the QinetiQ's approach found 12. The Envisat data set contained 11 oil slicks verified by the aircraft. Eight of these slicks were found by the KSAT operator, eight were found by the algorithm, and five were found by QinetiQ's oil spill system. Fig. 5 shows examples of verified oil slicks that were correctly classified.

A very interesting scene is from July 21, 2003. For this date, the same area in the Finnish sector of the Baltic Sea was covered by both Envisat and Radarsat images. Parts of this scene are shown in Fig. 6. The Radarsat image was taken at 1605 and the Envisat image at 1945. In the Radarsat image, the leftmost slick was reported as oil by KSAT, QinetiQ, and the algorithm, whereas the rightmost slick was not reported. Both these slicks have been verified as oil. In the Envisat image, the rightmost slick was reported by KSAT, the algorithm, and QinetiQ, whereas the leftmost slick was not reported. The leftmost slick appears larger and clearer in the Radarsat image. It is possible that the oil has partly evaporated during this time. The rightmost slick appears with good contrast and separated from the surrounding low-wind pattern in the Envisat image, but in the Radarsat image, the contrast is lower, and it can be mistaken to be part of the low-wind pattern. Note also that in the Envisat image, a ship or oil rig can be seen adjacent to the slicks. However, this is probably not the source of the slick, as the slick was also visible on the Radarsat image taken 3.5 h earlier.

In addition to the capabilities of detecting the verified oil slicks, we can also study the slicks reported by satellite but verified as not oil by the aircraft. For Envisat, nine slicks were detected by the algorithm; the location was checked by the aircraft but was found to contain no oil. Six of these nine slicks (all from the Baltic Sea) were verified as algae. Algae are a problem in the Baltic Sea because the type of algae observed here will dampen the capillary waves and will result in a dark spot in the SAR image. Additional information about algal activity is desirable in the Baltic during the summer. For two of the nine false alarms, nothing was visible when the aircraft checked the location a few hours later. In this case, the slicks were small, and the oil might have disappeared in the time between satellite image acquisition and aircraft inspection. An example of a slick verified as algae is given in Fig. 9.

We can also try to compare the number of alarms for the automatic algorithm to the operator. For the Radarsat data set, the operator produced 75 alarms, whereas the algorithm produced 71 alarms. Thus, they are expected to be comparable in terms of false alarm ratio. (Note here that all additional slicks reported by the algorithm could not be verified because the images were not processed in real time.) For the Envisat data set, the operator reported 69 slicks, whereas the algorithm reported 112 slicks.

In judging the performance of an automatic algorithm compared to manual detection, it is also interesting to study the variance in detections done by different operators at KSAT. In the benchmark study, two different operators KSAT-1 and KSAT-2 inspected the same SAR images without knowing the results of the other inspections. For Radarsat, the two operators at KSAT detected 75 and 68 oil slick candidates, whereas the algorithm detected 71 slick candidates. Note that not all these slicks could be verified by aircraft, as only KSAT-1 inspected the images in real time. Of the 18 slicks detected and verified by KSAT-1, KSAT-2 detected 15, and the algorithm detected 14. For Envisat, KSAT-1 detected 11 slicks that were verified by aircraft. KSAT-2 detected eight of these slicks, and the algorithm also detected eight slicks.

A question that arises is why some of the verified oil slicks were not detected by any of the approaches. Fig. 7 shows verified oil slicks from Radarsat images that were not detected. The two slicks in the upper row are low-contrast slicks in a locally homogeneous background, although the scene does contain many low-wind areas with many look-alikes. We expect that the algorithm did not classify these two slicks as oil because the wind estimate was too low due to the fact that the surroundings (in a larger local window) contained low-wind areas. Improving the local wind estimate will help this. In the second row of Fig. 7, the slicks are in very heterogeneous surroundings with complex look-alike patterns. It is very difficult to discriminate these slicks from the surroundings. Low-contrast slicks on a locally homogeneous background was found to be difficult to detect on Radarsat images, but this was not a problem for Envisat images (perhaps due to the more favorable VV polarization). The slicks not detected on Envisat images were those that overlapped with dark areas due to low wind (see Fig. 8 for examples).

For Radarsat images, the processing time for manual inspection at KSAT was between 3 and 25 min, with an average inspection time of 9 min. QinetiQ's semiautomatic approach took
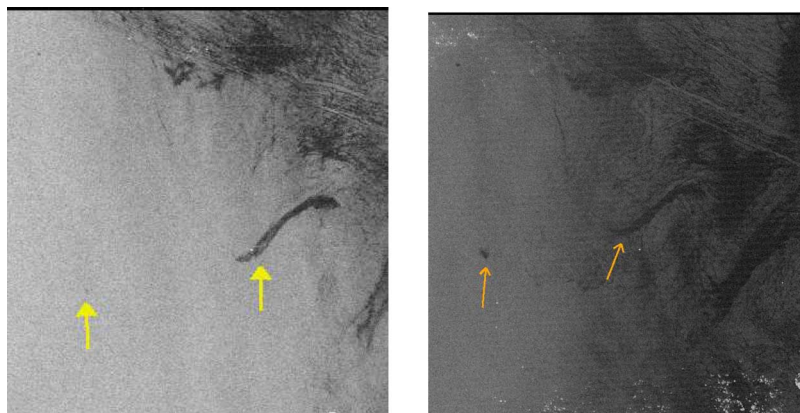
Fig. 6.   (Left) Envisat and (right) Radarsat image of the same scene taken on the same date. On the Envisat image, the leftmost spill was missed, whereas the rightmost spill was correctly classified as oil. In the Radarsat image, the result was the opposite (copyright ESA/Canadian Space Center/KSAT/NR).
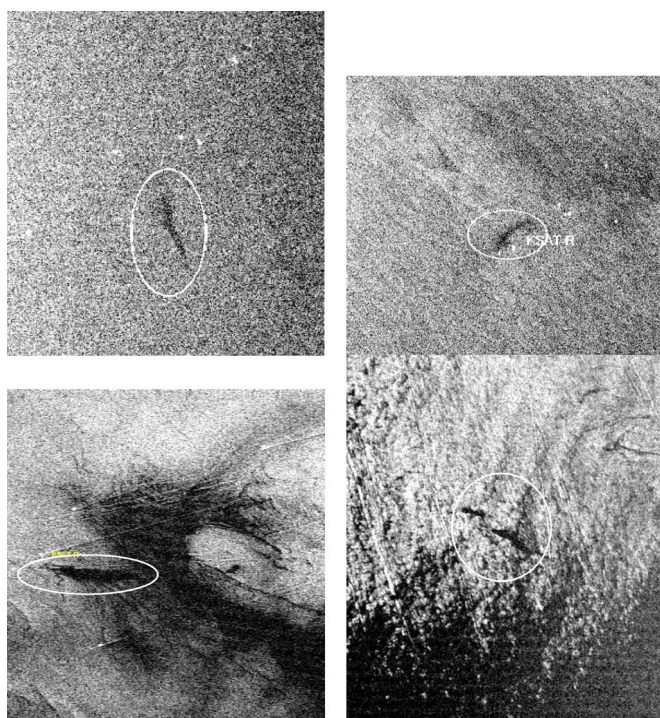

Fig. 7.   Examples of oil slicks verified by the aircraft but not detected by the operator or the algorithm (copyright ESA/KSAT/NR).
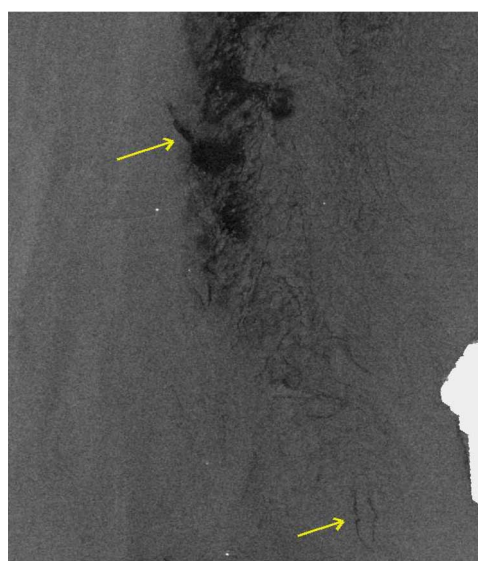

Fig. 8.   Examples of oil slicks from September 2, 2003 (Envisat), verified by the aircraft but not detected by the operator or the algorithm (copyright ESA/KSAT/NR).

7–32 min per scene with an average of 20 min. The automatic algorithm had an average processing time of 3 min per Radarsat image. The processing time for Envisat images for KSAT was 10 min per scene on the average and varied between 5 and 23 min. QinetiQ used 6–29 min with an average of 18 min, whereas the automatic algorithm had an average processing time of 1.45 min on a 2-GHz Linux computer. Note that even if the scene size for Radarsat SCN is smaller than that for Envisat (300×300 km compared to 400×400 km), the resolution is higher for Radarsat; thus, the processing time is longer.

V. DISCUSSION AND CONCLUSION

In this paper, we have presented a system for automatic detection of oil spills in Radarsat and Envisat images. A major part of the oil spill detection problem is to distinguish oil slicks
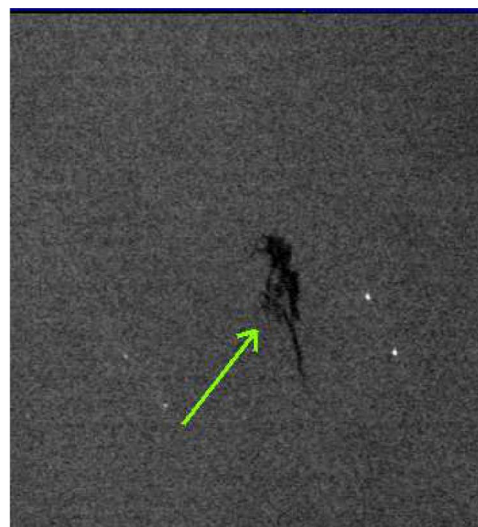

Fig. 9.   Slick from August 7, 2003, verified as algae by the aircraft (copyright ESA/KSAT/NR).

from other natural phenomena that create dark patches in the SAR image. The oil spill approach can be divided into detection of dark areas, extracting features for all dark areas, and then classifying all spots as oil spills or look-alikes. A multiscale approach for dark spot segmentation was presented. Sensor-specific modules for Radarsat and Envisat images have been developed. Slick feature extraction was based on features that describe slick shape, contrast, homogeneity, and surroundings. The classification approach is a mix between a statistical classifier with subclasses based on wind and shape and a rule-based approach. The rule-based approach was introduced to reduce the number of false alarms.

In some cases, algae can be misclassified as oil because certain algae types have a similar effect to oil in dampening the radar backscatter signal. Additional information about algal blooms, or the likelihood of observing algae in a given geographical location, should be utilized in the classification step. We also consider to include oil spill hot spot information in terms of a database with previously reported oil slicks and their location.

The dark spot segmentation step works well in most situations, but it can sometimes fragment a thin linear slick into several shorter segments. We are currently working on improving the segmentation for such slicks. Future work on this topic will also include a thorough evaluation on the use of different classifiers for the dark classification step.

The algorithm has been benchmarked against manual oil spill detection and a semiautomatic approach on 59 Radarsat and Envisat images. The benchmark showed that two manual operators do not identify exactly the same oil slicks. The performance of the automatic system is quite comparable to manual inspection, both in terms of performance in detecting verified oil spills and in terms of the number of alarms. The automatic algorithm is faster than manual inspection, with an average processing time of 1.45 min compared to 10 min for manual inspection.

In determining the best procedure for operational monitoring of oil spills, we find that the algorithm can be a very interesting supplement to manual inspection. Aircraft are still needed for oil spill verification. We propose to run the algorithm as part of an oil spill detection service by including a manual step prior to sending out the aircraft. This would improve the existing fully manual service, as the benchmark revealed that there is some variance between different operators. Currently, a prototype version of this algorithm has been implemented at KSAT.

REFERENCES

[1] C. Brekke and A. Solberg, "Oil spill detection by satellite remote sensing," *Remote Sens. Environ.*, vol. 95, no. 1, pp. 1–13, 2005.

[2] B. Fiscella, A. Giancaspro, F. Nirchio, P. Pavese, and P. Trivero, "Oil spill detection using marine SAR images," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3561–3566, 2000.

[3] F. Del Frate, A. Petrocchi, J. Lichtenegger, and G. Calabresi, "Neural networks for oil spill detection using ERS SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2282–2287, Sep. 2000.

[4] A. H. S. Solberg, G. Storvik, R. Solberg, and E. Volden, "Automatic detection of oil spills in ERS SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 4, pp. 1916–1924, Jul. 1999.

[5] T. F. N. Kanaa, E. Tonye, G. Mercier, V. Onana, J. Ngono, P. Frison, J. Rudant, and R. Garello, "Detection of oil slick signatures in SAR images by fusion of hysteresis thresholding responses," in *Proc. IGARSS*, 2003, vol. 4, pp. 2750–2752.

[6] L. Y. Change, K. Chen, C. Chen, and A. Chen, "A multiplayer–multiresolution approach to detection of oil slicks using ERS SAR image," in *Proc. 17th ACRS*, Colombo, Sri Lanka, 1996.

[7] C. F. Chen, K. S. Chen, L. Y. Chang, and A. J. Chen, "The use of satellite imagery for monitoring coastal environment in Taiwan," in *Proc. IGARSS*, 1997, vol. 3, pp. 1424–1426.

[8] A. K. Liu, C. Y. Peng, and S. Y.-S. Chang, "Wavelet analysis of satellite images for costal watch," *IEEE J. Ocean. Eng.*, vol. 22, no. 1, pp. 9–17, Jan. 1997.

[9] S. Y. Wu and A. K. Liu, "Towards an automated ocean feature detection, extraction and classification scheme for SAR imagery," *Int. J. Remote Sens.*, vol. 24, no. 5, pp. 935–951, 2003.

[10] M. Indregard, A. Solberg, and P. Clayton, "D2-report on benchmarking oil spill recognition approaches and best practice," 2004, Oceanides project, Eur. Commiss., Archive No. 04-10225-A-Doc, Contract No: EVK2-CT-2003-00177, Tech. Rep. [Online]. Available: http://oceanides.jrc.cec.eu.int

[11] M. Barni, M. Betti, and A. Mecocci, "A fuzzy approach to oil spill detection on SAR images," in *Proc. IGARSS*, 1995, vol. 1, pp. 157–159.

[12] A. Gasull, X. Fabregas, J. Jimenez, F. Marques, V. Moreno, and M. Herrero, "Oil spills detection in SAR images using mathematical morphology," in *Proc. EUSIPCO*, Toulouse, France, Sep. 2002, vol. 1, pp. 25–28.

[13] I. Keramitsoglou, C. Cartalis, and C. Kiranoudis, "Automatic identification of oil spills on satellite images," *Environ. Model. Softw.*, vol. 21, no. 5, pp. 640–652, 2006.

[14] H. A. Espedal and O. M. Johannessen, "Detection of oil spills near offshore installations using synthetic aperture radar (SAR)," *Int. J. Remote Sens.*, vol. 21, no. 11, pp. 2141–2144, 2000.

[15] T.-I. Bern, S. Moen, T. Wahl, T. Anderssen, R. Olsen, and J. A. Johannessen, "Oil spill detection using satellite based SAR. Completion report for phases 0 and 1," Trondheim, Norway, OCEANOR Tech. Rep. OCN-R92071, 1992.

[16] M. Perez-Marrodan, "ENVISYS—Environmental monitoring warning and emergency management system," in *Proc. AFCEA Kiev Semin.*, May 28–29, 1998, pp. 122–132.

[17] P. Wessel and W. H. F. Smith, "A global self-consistent, hierarchical, high-resolution shoreline database," *J. Geophys. Res.*, vol. 101, no. 1, pp. 8741–8743, 1996.

[18] M.-K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. Inf. Theory*, vol. IT-8, no. 2, pp. 179–187, Feb. 1962.

[19] L. Tufte, "D6(a)-report on oil spill data standardisation, oil spill pattern analysis and hot spots, as well as pollution conditions in test sites including oil spill maps and time series," 2004, Oceanides project, Eur. Commiss., Archive No. 04-10225-A-Doc, Contract No: EVK2-CT-2003-00177, Tech. Rep. [Online]. Available: http://oceanides.jrc.cec.eu.int

Anne H. S. Solberg (S'92–M'96) received the M.S. degree in computer science and the Ph.D. degree in image analysis from the University of Oslo, Oslo, Norway, in 1989 and 1995, respectively.

During 1991–1992, she was a Visiting Scholar at the Department of Computer Science, Michigan State University, East Lansing. She is currently with the Digital Signal Processing and Image Analysis Group, Department of Informatics, University of Oslo, as an Associate Professor. Her research interests include SAR image analysis, oil spill detection, hyperspectal imagery, statistical classification, and data fusion.

Camilla Brekke received the Cand.Sci. degree in computer science from the University of Oslo, Oslo, Norway, in 2001. She is currently working toward the Ph.D. degree in image analysis and remote sensing, in collaboration with the Digital Signal Processing and Image Analysis Group, Department of Informatics, University of Oslo.

From 2001 to 2002, she was a Young Graduate Trainee at the European Space Agency, European Space Research and Technology Center, Noordwijk, The Netherlands. She is currently with the Norwegian Defence Research Establishment, Kjeller, Norway, as a Scientist.

Per Ove Husøy received the B.Sc. degree in electrical engineering from Texas Tech University, Lubbock, in 1982 and the Dr.Ing. degree in electrical engineering from the Norwegian Institute of Technology, Trondheim, in 1991. His thesis work was on automatic speech recognition.

Since 1991, he has worked with the R&D departments of the Kongsberg Group, Siemens, Thales, and the Norwegian Computing Center in the area of signal processing and pattern classification. He is currently with the R&D Department, TANDBERG, Lysaker, Norway, as a Senior Engineer. His research interests include audio, video, and image processing and classification.