AUTOMATIC DETECTION AND ESTIMATING CONFIDENCE FOR OIL SPILL DETECTION IN SAR IMAGES

Anne Schistad Solberg[1]

[1]Department of informatics, University of Oslo, Norway
[*]Also at Norwegian Computing Center, Oslo, Norway

ABSTRACT

We present algorithms for automatic detection of oil spills in SAR images. The algorithm consists of three main parts, dark spot detection, spot feature extraction, and spot classification. The algorithms have been trained on a large number of ERS, RADARSAT and ENVISAT images. Our algorithm performed well both on RADARSAT and ENVISAT images, and can be considered a good alternative to manual inspection if large ocean areas are to be inspected. We also present methods for estimating the confidence automatically as part of the classification process. Confidence assignments are based on the computed features. The automatically assigned confidence estimates are compared to manual assignment of confidencel levels.

Key words: SAR, oil spill detection, ENVISAT.

1. INTRODUCTION

The potential for oil spill detection in SAR images was demonstrated on ERS images in the early 90-ies. Oil slicks on the sea surface are seen relatively often. Automatic identification of oil spills in SAR images is complex because of features that resamble oil spills, called look-alikes, which often occur in low wind conditions. The SAR signatur of an oil spill and its surroundings depends on a number of parameters like wind speed, wave height, and the amount and type of oil released. The shape of the spill will depend on whether the oil was released from a stationary object or from a moving ship, the amount of oil involved, and the wind and current history between the release and the image acquisition. A trained human operator is mostly able to discriminate between oil slicks and look-alikes based on experience, extracted information like location and weather conditions, and by considering shape and contrast between the feature and the surrounding sea. Such considerations have to be incorporated into an automatic classification system as well.

We started to work with automatic oil spill detection in ERS SAR images in 1993, and developed an advanced algorithm for oil spill detection, which performed well on alarge-scale study involving 80 ERS SAR images [Solberg et al. 1999]. This framework has now been extended to RADARSAT and ENVISAT ASAR images, by adding sensor specific modules and traning the system on a high number of images from each sensor. The three modules use very similar features, but the classification step has a sensor-specific set of rules.

Oil spill detection based on manual inspection of SAR images is today used in combination with surveillance aircraft flights in many European countries. The suspect slicks are assigned confidence levels low, medium and high and this information is transmitted to the surveillance aircraft in near-real time, so that the aircrafts can check the reported positions and possibly catch the polluter. A study performed by the European Union project Oceanides investigated the performance of detection based on satellite imagery compared to aircraft detection [Indregard et al. 2004]. This study showed that the conficence assigned by the operator would be of high value to the aircraft if it is reliable. However, experiments involving several operators showed that reliable and consistent assignment of confidence levels is difficult and the procedure is subjective.

In this paper, we present methods for estimating this confidence automatically as part of the classification process. Confidence assignments are based on the computed features, and the procedure is objective. The performance is compared to two manual operators on a benchmark data set.

2. METHODOLOGY FOR OIL SPILL DETECTION

Oil slicks dampen the Bragg waves on the ocean surface and reduce the radar backscatter coefficient. This results in dark regions or spots in a satellite SAR image. A part of the oil spill detection problem is to distinguish oil slicks from other natural phenomena that dampen the short waves and create dark patches on the surface. Look-alikes may include natural film, grease ice, threshold wind speed areas (wind speed < 3m/s), wind sheltering by land, rain cells, shear zones, internal waves and natural oil seepage [Espedal 1998].

Our oil spill algorithms consists of three main parts, detection of dark spots in the image, computing features from these regions, and classifying each region as either oil spill or look-alike. Several other published papers [Fiscella et al. 2000, Frate et al. 2000] on oil spills detection follows this framework, which we introduced in 1996 [Solberg et al. 1996].

The algorithm includes preprocessing, masking of land areas, detection of dark spots, spot feature extraction and dark spot classification. The first step converts the original SAR data to a common format and geographical projection. A land mask is applied in order to mask away all land areas.

RADARSAT ScanSAR images are divided into strip lines in the range direction in order to optimize processing time. This also prohibits large variations in backscatter across the relatively large incidence angle range.

For ENVISAT ASAR Wideswath images, preprocessing starts with land masking the images by converting a land mask to the SAR image grid to avoid resampling the speckle pattern. A normalization of the backscatter with respect to incidence angle is done. A model for incidence angle correction was compared to an approach based on fitting a flat profile to the range profiles, and the profile fitting approach performed the best and was thus selected for further processing.

2.1. Dark spot detection

The algorithm for detecting dark spots is based on adaptive thresholding. This thresholding is based on an estimate of the typical backscatter level in a large window. The adaptive threshold is set to k dB below the estimated local mean backscatter level. Wind data is used to determine k.

For RADARSAT ScanSAR images, a pyramid algorithm was needed. An image pyramid where a pixel on level N consists of the mean of M pixels at level N-1 is constructed. Each level in the pyramid is thresholded yielding a binary pyramid. The final spot image is constructed by combining the different binary images in the pyramid.

Using three levels in the pyramid did not work well for ENVISAT, due to the smaller pixel resolution of the selected product types. Thus, a two-level pyramid is used for ENVISAT. If no wind information is available as input, the local homogeneity is used to compute the threshold in the following manner. For SAR image, the power-to-mean ratio (PMR) is a good measure of homogeneity. Scenes with high or medium wind are expected to have a small number of oil spill look-alikes, and the sea will be fairly homogeneous as measured by the PMR-ratio. In low-wind conditions, a high number of look-alikes is often observed, although local variations are common. By computing the PMR-rato in local windows, we get an indication about the number of look-alikes in the scene. The threshold is adjusted locally based on the PMR-ratio.

2.2. Slick feature extraction

Both the RADARSAT and the ENVISAT modules use the same set of features (however with slight modifications, in particular for the distance to bright spot operator, which contains a simple ship detection algorithm tailored to each sensor.) This set of features is computed for each detected spot of a certain minimum size. The features are a mix of standard region descriptors from image analysis, and features tailored to oil spill detection. The features are [Solberg et al. 1999]:

- Slick complexity
- Slick power-to-mean ratio
- Slick local contrast
- Slick width
- Slick local neighbours
- Slick global neighbours
- Border gradient
- Slick area
- Distance to detected ship
- Slick planar moment
- Number of regions in the image
- Slick smoothness contrast

2.3. Slick classification

After spot detection and feature extraction, we have a set of M dark spot objects that we want to classify as either oil slick or look-alike. The classification framework used for ERS, RADARSAT and ENVISAT images is a combination of a statistical classifier model which incorporates prior knowledge in terms of loss functions, and a rule-based approach (see [Solberg et al. 1999] for details). Prior probabilities for oil or look-alike is adjusted based the rule-based corrections. Slicks are divided into different subgroups based on wind level and shape as measured by the planar moment. For each subgroup, a class description base for oil and look-alikes contains the parameters for a statistical classifier (Gaussian).

The output from the classifier is an image containing only the slicks classified as oil, in addition to their corresponding geographical co-ordinates and a confidence estimate.

2.4. Oil spill confidence assignment

The assigned confidence will be compared to the confidence assigned by trained operators at Kongsberg Satellite Service (KSAT) as part of their oil spill detection service.

2.4.1. Manual spill confidence assignment

Kongberg Satellite Services have designed at set of rules used by their operators in assigning oil spill confidence [Indregard et al. 2004]. The opeators at KSAT have been trained to analyse SAR images for detection of oil pollution. To confirm the reliability of a detected possible oil pollution the operators utilize information about wind speed and direction, oil rig/pipeline locations, national territory borders and coastlines. After a pollution is detected it is classified as high, medium or low confidence of being actual pollution. The classification can also give other attributes to the detected pollution like possible name and location of source, orientation and other remarks.

The following guidelines are used to determine the confidence level:

High confidence:

- The slick has a large contrast to the surroundings.
- The surrounidings are homogeneous.
- The wind speed is moderator to high, i.e. approximately 6-10 m/s.
- Ship or platform directly connected to slick.

Medium confidence:

- The wind speed is moderate to low, i.e. approximately 3-6 m/s.
- The slick has a diffuse/low contrast to the graylevel surrounding in moderate to high wind speeds.
- The shape of the slick is irregular, i.e. the edges are not smooth.

Low confidence:

- Low wind areas are located nearby.
- Natural slicks (e.g. biological algae or fractal streaks at very low wind) are located nearby.
- The slick has diffuse edges and/or an irregular shape.

All of the different aspects are assessed together and against each other to determine the confidence level. All results will have some degree of uncertainty in them, and experienced operators are an important factor.

2.4.2. Automatic spill confidence assignment

The first attempt on computing automatical confidence levels was based on directly translating the criterias used at KSAT to computed features. Homogeneity of the surrounding was estimated using the PMR feature and the number of neighboring dark areas, contrast to the surroundings using the slick local contrast feature, and the distance to ship or platform using the distance to point source feature, and with a very rough estimate of wind levels. However, using these criteria, almost all slicks were assigned low confidence, and a new procedure had to be defined.

The simple rules established by KSAT were extended to rules involving more of the features. Local contrast is now computed using both the features SLICK LOCAL CONTRAST and BORDER GRADIENT. In addition, the shape of a slick was found important as lookalikes have more complex shapes. Linear slicks have higher probability of being oil. The linearity is computing using the SLICK PLANAR MOMENT feature. By using the training data set, the set of rules was extended to approcimately 20 rules assigning oil slick confidence levels from low to high. This was done by looking at the slicks classified as oil in the traing data set, and studying the actual computed values for the different features. The performance of these rules is then evaluated on a benchmark data set and compared to the assignment given by manual operators at KSAT.

2.5. Experimental results

The classifier has been trained on approximately 100 RADARSAT images and 64 ENVISAT images from European waters (mainly the Baltic and the North Sea). All the SAR images were processed by KSAT. The training data set was made available to us by KSAT as part of the EC project Oceanides. All the training images contained oil spill candidates. For some of these images, the reports sent from KSAT to the surveillance aircraft were available to us, but for more than half of the images, such a report was not available.

Training an automatic oil spill detection algorithm can be done with a certain false alarm ratio in mind. The algorithm can be tuned to produce an alarm for all suspicious slicks in a scene, or to produce alarms only for slicks with a large probability of being oil. Automatic oil spill detection will often be followed by manual inspection/verification, before an alarm is sent to a surveillance aircraft. To be able to detect all true oil spills, a certain number of false positives must be expected. In this study, training was done by manually inspecting all SAR images and marking all suspect slicks as oil. A doubt class was used to avoid that doubt cases were included in the look-alike class, as all slicks not marked as oil or doubt are used to train the look-alike class. Slicks are divided into different subclasses automatically based on shape and wind level, so that during classification a slick is matched with the corresponding subclass given wind level and shape as computed from the first planar moment. All slicks with conficence low, medium or high as computed by the algorithm are reported as oil slick candidates.

2.6. Benchmarking results

In addition to the training data, benchmark images with aircraft verifications were available to test the performance. As part of the EC project Oceanides, a joint satellite-airborne campaign was done 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 Finish and German sectors of the Baltic sea, in addition to the German sector of the North Sea. A total of 60 RADARSAT and ENVISAT images were acquired between July and December 2003.

The RADARSAT benchmark data set consists of 32 RADARSAT images, and the ENVISAT benchmark data set of 28 ENVISAT ASAR images. They were analysed by an operator at KSAT, by a semi-manual algorithm developed at QinetiQ, and by our algorithm [Indregard et al. 2004]. 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 performance of the three approaches was compared both in terms of capabilities in detecting verified oil spills and the number of oil spills reported. Verification of all slicks found in the images was not possible, as many of the slicks were outside the Finnish or German territories (meaning that the aircraft was not allowed to fly to the location), or because flight plans or satellite acquisitions sometimes had to be cancelled.

The RADARSAT data set contained 18 oil slicks verified by the aircraft. The operator at KSAT found 15 of these slicks, our algorithms found 14, while QinetiQ's semi-automatic approach found 12.

The ENVISAT data set contained 8 oil slicks verified by the aircraft. The operator at KSAT found 7 of these slicks, our algorithms found 7, while QinetiQ's semi-automatic approach found 5.

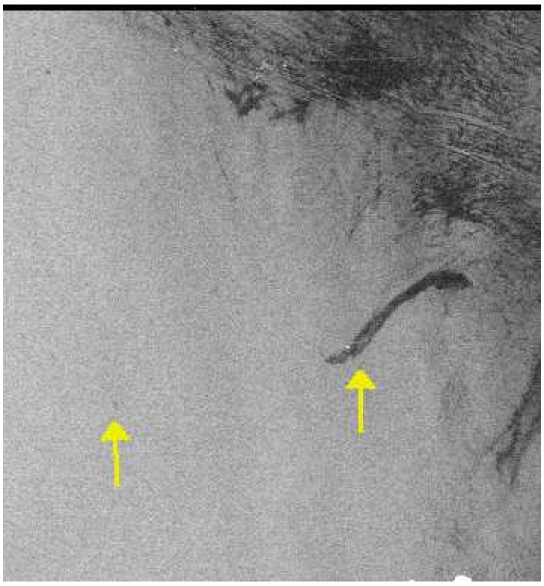
Slick number	KSAT-1	KSAT-2	Algorithm
1	Medium	Medium	High
2	Medium	Medium	Medium
3	Medium	Low	Medium
4	High	Medium	Medium
5	Medium	Medium	Low
6	High	-	Medium
7	High	-	High
8	High	Low	Low
9	Medium	Low	Medium
10	Medium	-	High
11	High	Low	High
12	High	Low	High
13	High	High	High
14	High	Low	High
15	High	Low	Low
16	High	-	Medium
17	High	-	High
18	High	Medium	High
19	High	High	High
20	High	High	High
21	High	High	High
22	High	Low	Low

Table 1. Comparison of confidence levels assigned by two different operators at KSAT and the automatically assigned confidence levels.

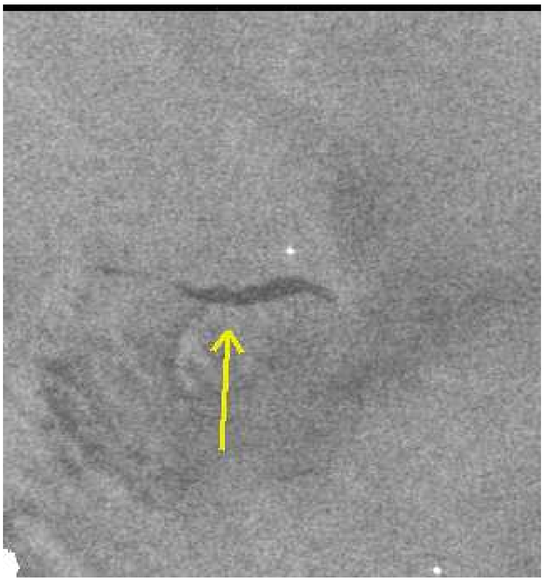
Figure 1 shows examples of verified oil slicks correctly classified, and their corresponding confidence levels associated by two different operators at KSAT, and the automatic algorithm.

Table 1 presents a table of oil slick candidates reported by both KSAT and the automatic algorithm, and their associated confidence. As we can see, there is significant variation in the reported confidence. The two manual operators agrees for 7 of 22 slicks, have a difference of 1 level for 4 slicks, and a difference of 2 levels for 6 slicks. The best agreement is found between KSAT-1 and the automatic algorithm, which agree for 13 of 22 slicks, differ with 1 level for 6 slicks, and differ with two levels for 3 slicks. In general KSAT-2 reports lower confidence than the others, with median confidence level low, while KSAT-1 and AUTO have median confidence level high. KSAT-1 did not report any low confidence slicks.

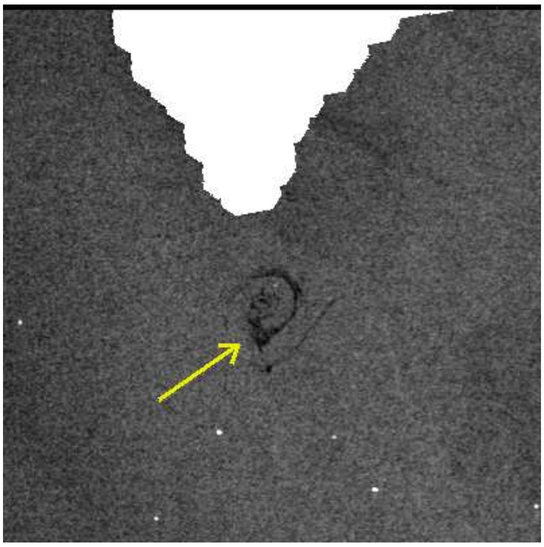
Figure 2 shows a slick reported as oil, but verified as algae. In certain areas, the SAR sensor is not sufficient to discriminate between oil and natural films or algae, particularly in the Baltic Sea where algae is common during the summer. Linear slicks on homogeneous background are suspect in nature, and should be inspected, although they can sometimes be mixed with algae. Using additional information about algal blooms should be possible and will be tested at a later stage.



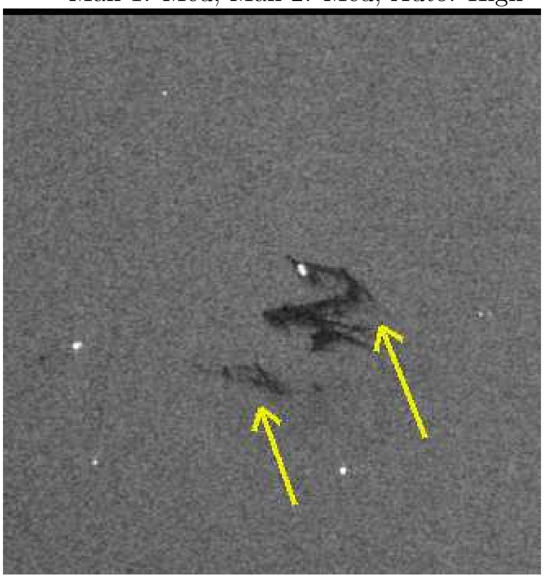
Man-1: Low, Man-2: Med, Auto: High



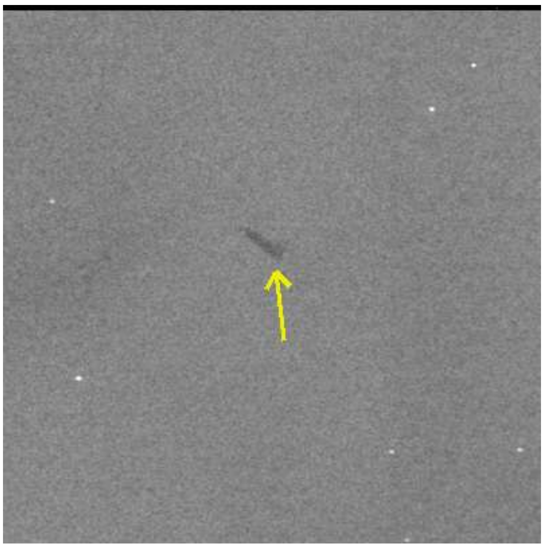
Man-1: Med, Man-2: Med, Auto: High



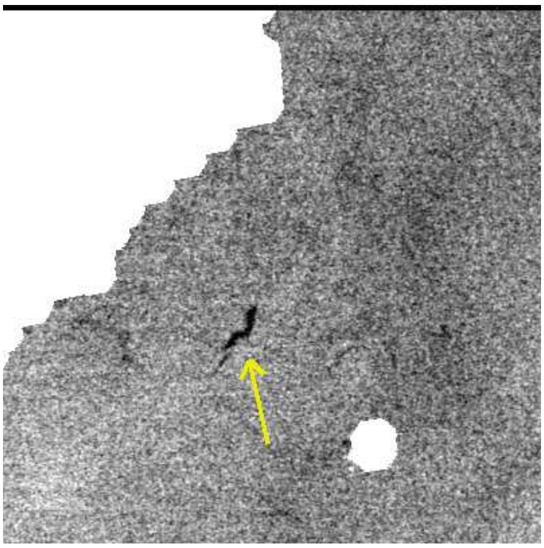
Man-1: High, Man-2: Low, Auto: Low



Man-1: High, Man-2: High, Auto: High



Man-1: High, Man-2: High, Auto: High



Man-1: High, Man-2: Low, Auto: Low

Figure 1. ENVISAT images with examples of verified oil slicks. The confidence levels assigned by two different operators at KSAT are given, together with the confidence assigned by the automatic algorithm.©ESA/KSAT/NR

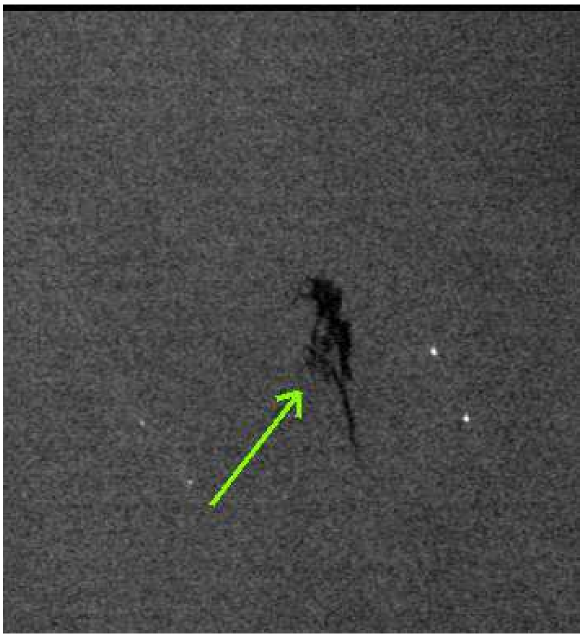


Figure 2. ENVISAT image with one suspicious slick repored by satellite, but verified as algae by the aircraft. ©ESA/KSAT/NR

2.7. Performance of confidence assignments

3. CONCLUSION

In this paper we have presented algorithms for detecting oil spills in RADARSAT and ENVISAT SAR images. A benchmark test showed that the algorithms performed well both on RADARSAT and ENVISAT SAR images, but that in some cases additional information was needed to distinguish between oil spills and natural films/algae. Future work on this topic will incorporate algae information into the algorithms, and compute improved confidence estimates for the detected oil slicks. A study of confidence levels assigned by two operators showed large deviations between operators. The proposed procedure for confidence assignment shows some potential, but we still believe that it can be improved to be more robust. Presently, a good practise might be to combine confidence assignement from the algorithm with a manually assigned confidence level.

As a consequence of the study, Kongsberg Satellite Services are now working to improve their training of the operators in assigning oil spill confidence levels. The automatic algorithm described in this paper has now been installed at KSAT to allow combined use.

ACKNOWLEDGMENTS

This work was partly funded by the EC project Oceanides. The author would like to thank the Oceanides project team, in particular Marte Indregard, KSAT and Peter Clayton, QinetiQ, for sharing the benchmark results.

REFERENCES

[Espedal 1998] H. Espedal, 'Detection of oil spill and natural film in the marine environment by spaceborne SAR', Ph.D. thesis, Department of Physics, University of Bergen, 1998.

[Fiscella et al. 2000] B. Fiscella, A. Giancaspro, F. Nirschio, P. Pavese and P. Trivero, 'Oil spill detection using marine SAR images', Int. J. Remote Sensing, vol. 21, pp. 3561-3566, 2000.

[Frate et al. 2000] F. D. Frate, A, Petrocchi, J. Lichtenegger and G. Calabresi, 'Neural networks for oil spill detection using ERS SAR data', IEEE Tr. Geosc. Rem. Sens., vol 38, pp. 2282-2287, 2000.

[Indregard et al. 2004] M. Indregard, A. Solberg and P. Clayton, *D2 - Report on benchmarking oil spill recognition approaches and best practice*, Deliverable D2, EU Contract EVK2-CT-2003-00177.

[Solberg et al. 1999] A. S. Solberg, G. Storvik, R. Solberg and E. Volden, *Automatic detection of oil spills in ERS SAR images*, IEEE Trans. Geosc. Rem. Sens, vol. 37, pp: 1916-1924, 1999.

[Solberg et al. 1996] A. S. Solberg and R. Solberg, 'A large-scale evaolution of features for automatic detection of oil spills in ERS SAR images', IGARSS-96, Lincoln, NE, May 1996, pp. 1484-1486.