Oil Spill Detection: Feature Extraction

Luca Riccardi, Vito Domenico Tagliente, Pietro Tota

January 2016

1 Abstract

This paper addresses oil spill detection from SAR images. SAR observations led to a significant improvement in sea oil slick observation since they allow distinguishing oil slicks from a broad class of lookalikes. In particular, it focuses on the automatic features extraction of regions of interest (ROIs) in two classes, namely oil spills or look-alikes. ROIs have been manually selected from SAR images. From each ROIs, a set of features has been extracted. Three kinds of features were indentified: Geometrical, Backscatter and Texture features. In addition, a fourth group related to the image was considerated: ancillary features. Then, these features have been used to populate a Dataset.

2 Introduction

It is worldwide recognized that a large amount of oil is yearly discharged into the marine environment due to vessel operations, land-based sources, etc. and that coastal environments suffer from petroleum pollutants. The topic is very relevant and it includes the different sources of sea oil spills, how they are affected by weathering and other physical processes, i.e. evaporation, dispersion, emulsification, and what are the effects of oil discharged into the sea on marine ecosystems. Oil sources can be classified into four categories: natural oil seeps, petroleum extraction, petroleum transportation, and petroleum consumption. The different tools used to detect and monitor oil spills are vessels, airplanes, and satellites. Vessels, especially if equipped with specialised radars, can detect oil at sea but they can cover a very limited area. The vessel, however, remains necessary in case oil sampling is required. The main systems to monitor sea-based oil pollution are the use of airplanes and satellites equipped with Synthetic Aperture Radar (SAR). SAR is an active microwave sensor, which captures two dimensional images. The brightness of the captured image is a reflection of the properties of the target-surface. The possibility of detecting an oil spill in a SAR image relies on the fact that the oil film decreases the backscattering of the sea surface resulting in a dark formation that contrasts with the brightness of the surrounding spill-free sea. Spaceborne SAR sensors are extensively used for the detection of oil spills in the marine environment, as they are independent from sun light, they are not affected by cloudiness, they cover large areas and are more cost-effective than air patrolling. Usually, for oil spill detection, large swath widths are chosen at the expense of lower resolution.

This approach is adopted because it is in our interest to cover as much area as possible even if very small oil spills can not be detected.
In the following sections we are going to give a brief overview about the features extracted then, how we worked about features extraction from setting workspace to dataset's creation.[1][2]

2.1 SAR imaging of oil spills

Oil films decrease the backscattering of the sea surface resulting in a dark formation on SAR images.

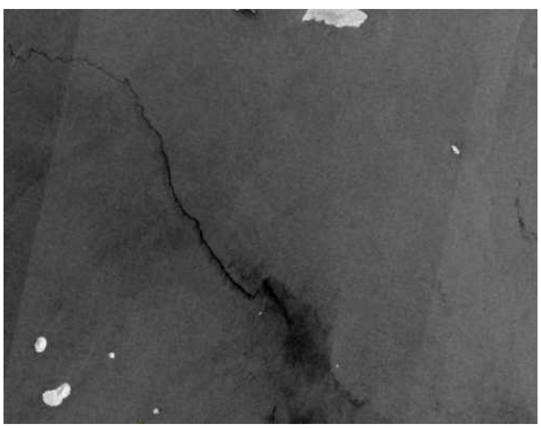


Figure 1: Oil dark patch in the sea

Although SAR-based solutions are really powerful for oil spill detection from space, mainly due to their all-weather and all-day detection capabilities, they have well known limitations. For example, the difficulty in detecting oil spills for high wind speeds ($\geq$ 6 m/s), which exposes to the possibility of missing true oil spills, or, when the wind speed is low ($\leq$ 1.5 m/s), to that of generating too many false alarms, i.e. alerts indicating supposed oil spills which could reveal to be look-alikes, after in site verification. Since the latter is very expensive, there is a need for keeping the number of false alarms as small as possible, while trying not to miss any true oil spill case. Thus, there is a need for alternative or complementary solutions: including some data from other source. In our project we are not going to include them, but they could be part of a future research.[4]

3 Features Extracted

An attempt was made to extract as many features as possible from those referred in the state-of-the-art literature. We collected 31 features.[3] These features have all been extracted from the segmented dark patch, except for the coast information, which was overtaken from the ancillary data or from the operators reports [1]. The used features can be divided in four main groups: geometrical, backscatter, texture and ancillary data features.

3.1 Geometrical features

1. Area (A).
2. Perimeter (P).
3. Complexity (C), defined as

$$C := \frac{P}{2\sqrt{\pi A}}$$

   This feature will take a small numerical value for regions with simple geometry and larger values for complex geometrical regions.

4. Length (L) = Ellipse-Length (EL): value of main axe of an ellipse fitted to the data.

[1] www.gnome.orr.noaa.gov

5. Width (W): approximate by the ratio of slick area to length.
6. Length To Width Ratio (LWR), defined as

$$LWR := \frac{L}{W}$$

7. Compactness (Comp), defined as

$$Comp := \frac{LWR}{A}$$

8. Ellipse-Width (EW): value of minor axe of an ellipse fitted to the data.
9. Ellipse-Asymetry (EA), defined as

$$EA := 1 - \frac{EW}{EL}$$

10. First Invariant Planar Moment (FIPM), defined as

$$FIPM := \frac{\sum_{i=1}^{n}[(x_i - x_c)^2 + (y_i - y_c)^2]}{n^2}$$

with

$x_c : mean\ of\ x\ and\ y_c :\ mean\ of\ y$

3.2 Backscatter

1. Inside Slick Standard Deviation (σ obj)
2. Inside Slick Radar Backcatter (μobj)
3. Outside Slick Standard Deviation (σsce)
4. Outside Slick Radar Backscatter (μsce)
5. Intensity Ratio ($\frac{\mu obj}{\mu sce}$)
6. Intensity Standard Deviation Ratio ($\frac{\sigma obj}{\sigma sce}$)
7. Intensity Standard Deviation Ratio Inside (ISRI), defined as

$$ISRI := \frac{\mu obj}{\sigma obj}$$

8. Intensity Standard Deviation Ratio Outside (ISRO), defined as

$$ISRO := \frac{\mu sce}{\sigma sce}$$

9. ISRI ISRO Ratio (IIR)
10. Mean Contrast (ConMe): difference (in dB) between the background mean value and the object mean value, defined as

$$ConMe := \mu sce - \mu obj$$

11. Max Gradient (GMax): maximum value (in dB) of border gradient magnitude, calculated using Sobel operator.
12. Mean Gradient (GMe): mean border gradient magnitude (in dB).
13. Gradient Standard Deviation (GSd): standard deviation (in dB) of the border gradient magnitudes.

3.3 Texture

Texture is a combination of repeated patterns with a regular frequency and texture analysis has often proved to be effective for oil spill classification. We have used gray level co-occurrence matrices (GLCM) to specify texture measures and have used the following measures:

1. GLCM Homogeneity

2. GLCM Contrast
3. GLCM Entropy
4. GLCM Correlation
5. GLCM Dissimilarity

3.4 Ancillary

This group, of so-called context features, includes features like:

1. Number of bright spots nearby (NBS)
2. Min distance to next bright spot (vessel, platform) (MDB)
3. Min distance from land (MDL)

Many of these extracted features are related to the same slick characteristics and are correlated. For example the features “Spreading”, “Compactness” and “LengthToWidthRatio”, both describe how long and thin objects are. In the same way, the features “First Invariant Planar Moment” and “Ellipse Assimetry” can all be considered to give an indication to the general shape of the object. In our work we wanted to examine as much of the features referred to in the literature as possible.

4 Our Approch

We tried our best to create a semi-automatic matlab script that allowed to extract all the features and fill a dataset with them. The entry task was split in many subtasks (Figure 6). Each of them aims to extract a group of features: Geometry, Backscatter, Texture and Ancillary.

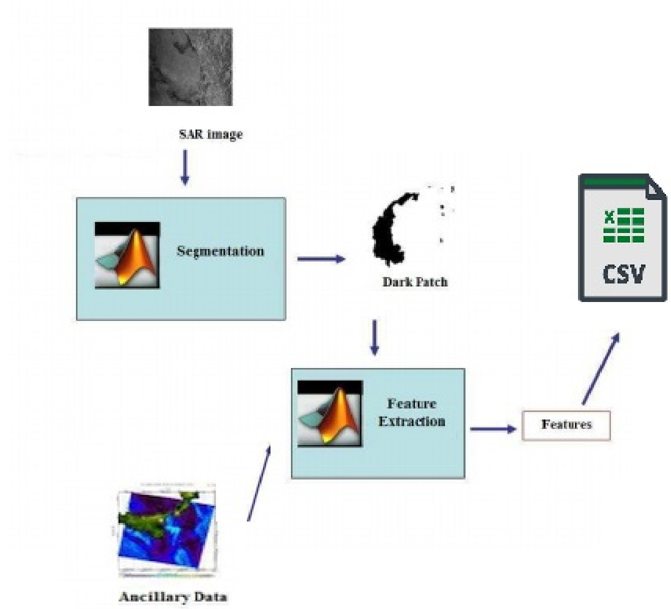
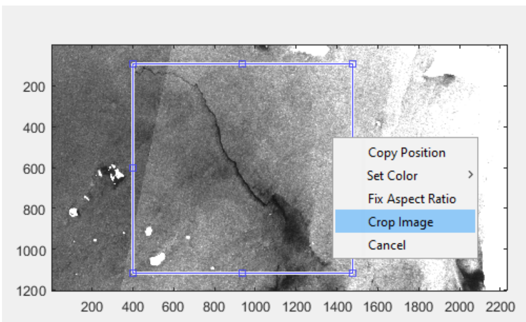


Figure 2: A panoramic of workflow

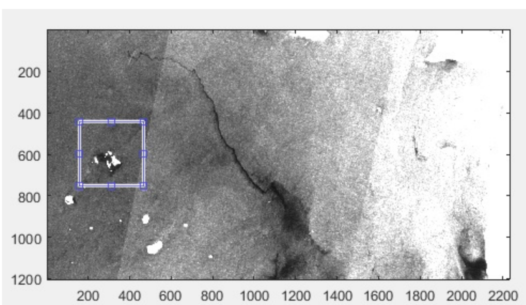
This choice was made to ensure the maintenance and modularity of the project. The test have been done using two SAR images: one for oil spill and the second for the look-alike one. About dataset we chose csv format, because it is the most popular and compatible format for data representation and portability.

4.1 Setting workspace

Before working with Matlab we needed to set the workspace. This means that we had to define a phase which was used to load images and do some preprocessing to get the workspace ready for the features extraction, which is the aim of this work. As for data, we used some geotiff images of an oil spill detected in Guimaras Island, Philippine, acquired from SAR telescope. We designed a script that can load data, crop the images using an area manually defined by operator and do some other tasks. We used a manual approach to select the local area, as you can see in figure 3, meanwhile the phase of features extraction implements the automatic approach of the methodologies used to detect oil spills.



(a) oil spill

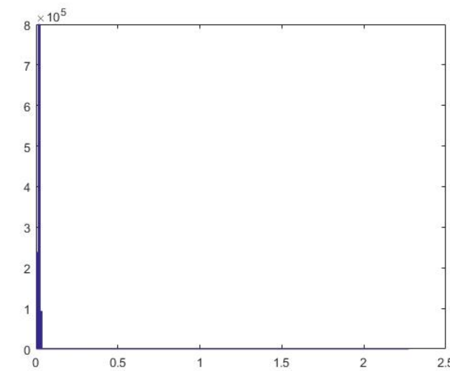


(b) look-alike

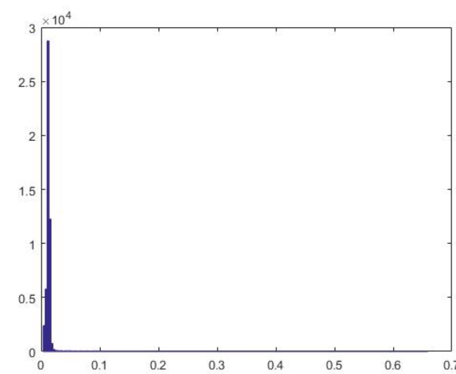
Figure 3: shows how to manually crop an image using the Matlab's tool

During this phase some preprocessing operations are done. This means that binary masks which have been used to manipulate spill or background pixels are generated. Conversion of data from sigma to dB representation is performed applying some filters to the SAR images. So at the end of this phase we obtain grayscale and dB representation of data, that will be used for the next phases.

Using the same images, we selected two areas: the first one containing the oil spill and the second one containing a look alike dark patch. We considered to compare the sigma and dB data representation histogram of these two patches, as shown in figure 4 and figure 5.



(a) oil spill histogram

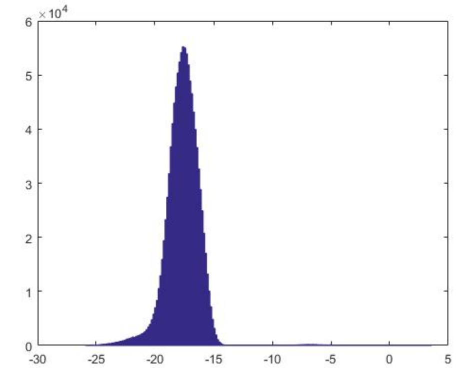


(b) look-alike histogram

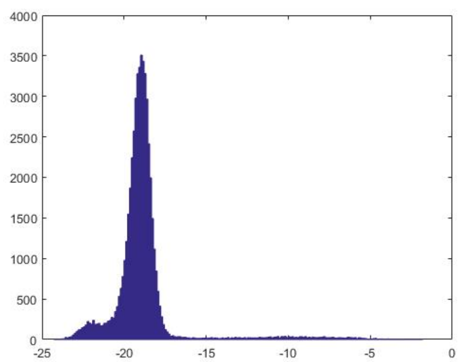
Figure 4: Sigma data representation

4.2 Feature extraction

We designed an hierarchy of scripts, in which each of them is able to perform a particular computation. For example, scripts like



(a) oil spill histogram



(b) look-alike histogram

Figure 5: dB data representation

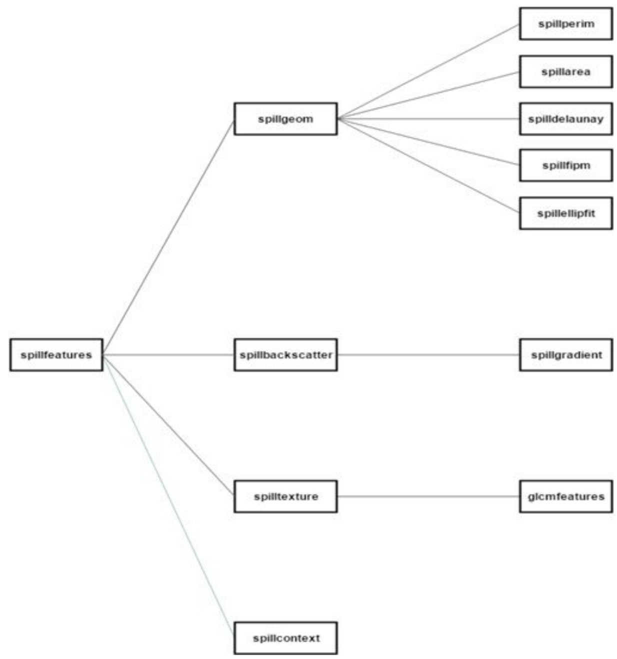


Figure 6: Hierarchy of scripts

”spillperim”, ”spillarea”, etc.. are used to process Geometrical analysis on the given spill patch data. The other scripts, on the other hand, are used to process Texture, Backscatter and Ancillary analysis. As regards the Geometrical features extraction, we processed the data starting from a grayscale representation.

We extracted the oil’s perimeter, the area, the complexity and the other stuff starting from the application of the masks on data, which were generated during the preprocessing phase.

Backscatter features were generated from dB representation of data, which is the real data acquired from the SAR telescope. Texture analysis were performed working on two types of data, the pixels of the spill patch and the background. In conclusion, we processed the Ancillary analysis. At this point, we dealt with the detection of the relationships between the oil spill and the context, the background data. This information is important for the classification analysis. For example, the light spots near the oil patch may represent the presence of a boat near the oil spill. All the work consists in detecting the key point in the image (f.e. oil slick barycenter, spikes barycenter, oil slick perimeter points etc. see Figure 7).

We used all the basic characteristics, which were processed in previous analysis, and we

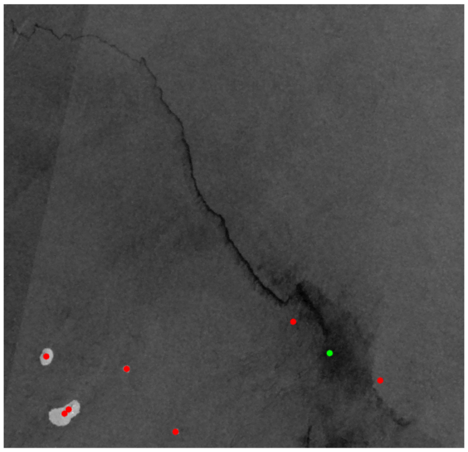


Figure 7: Oil slick (green) and spikes (red) barycenter identified by our Matlab script

correlated them to get these new ones. We found information about light spots while processing all the available data. On the other hand, to compute the distance of the oil spill from the coast we had to use a web service called GOODS [2] to obtain the coast data. GOODS generates BNA format files in which are stored data about the costs of the defined geographical area. We used this information to precess the minimum distance of the oil spill from the coast. To perform this operation we approximated the area of the oil spill as a single point, the oil slick barycenter.

4.3 Data extracted

In this section we can give a look to extracted data using the script from ROIs. We can compare data extracted from an oil spill and data extracted from a look-alike patch.

[2]www.gnome.orr.noaa.gov

Feature	Oil	Look-alike
A	9105.000000	1633.000000
P	1080830.750000	33826.375000
C	2.470564	2.504687
L/EA	1754.957323	307.813727
W	615.872954	109.892354
LWR	2.849544	2.801048
Comp	1.000000	1.000000
EW	1.154701	1.154701
EA	0.999342	0.996249
FIPM	0.007512	0.030286
σ obj	1.419728	3.429655
μobj	-0.263848	-1.004815
σsce	1.845056	3.823762
μsce	-17.456341	-17.831821
Intensity Ratio	0.015115	0.056350
ISDR	0.769477	0.896932
ISRI	-0.185844	-0.292978
ISRO	-9.461144	-4.663424
IIR	0.019643	0.062825
ConMe	-17.192493	-16.827007
GMax	40.719727	40.384918
GMe	32.364525	34.100403
GSd	31.019098	30.511026
GLCM Homogeneity	0.998415	1.000000
GLCM Contrast	0.003170	0.000000
GLCM Entropy	0.027708	0.000000
GLCM Correlation	0.228736	0.000000
GLCM Dissimilarity	0.003170	0.000000
NBS	8.000000	14.000000
MDB	116.851438	7.865683
MDL	43.285263	303.251171

4.4 Dataset creation

About this section we can repeat that we chose CSV format to store data, because it is the most popular and compatible format for data representation and portability. It's creation is totally automatic. It means that we can define a folder, which contains some images and automatically process them all. All the features described before are extracted for each of these images and the given data are saved into a CSV file.

5 Conclusion and Future work

Our system is semiautomatic, this means that during the first phase it requires an user input to define the crop area of the oil spill. This operation is performed using an user friendly graphic interface. After that manual operation, all the other work is completely automatic, which means that our system extracts all the features from the oil spill. Our system can also work with a dataset of geographical images. In so doing we save all the data of all computed images into a csv format file. A possible future work on the system is to implement an automatic oil/look-alike cropping procedure or, at least, give the user a choice between using automatic or manual crop method. This is why another expansion of this work can be the design of a binary classifier. For instance a classifier is a program that, after being "trained" with a training set of data, can choose a class for any new instance with a certain accuracy. In our case that would be a binary classifier because of the nature of the problem (there are only two possible class for any image: oil and look-alike).

References

[1] F. Nunziata A. Buono M. Migliaccio. Sar polarimetry for sea oil slick observation. *International Journal of Remote Sensing*, 2015.

[2] Marco Cococcioni Linda Corucci Andrea Masini Fabio Nardelli. Svme: an ensemble of support vector machines for detecting oil spills from full resolution modis images. 2011.

[3] Sonia Antunes Pelizzari. Oil spill detection using sar images. 2011.

[4] Konstantinos N. Topouzelis. Oil spill detection by sar images: Dark formation detection, feature extraction and classification algorithms. 2008.