Detection of Anomalous Particles from the Deepwater Horizon Oil Spill Using the SIPPER3 Underwater Imaging Platform

Sergiy Fefilatyev[1], Kurt Kramer[1], Lawrence Hall[1], Dmitry Goldgof[1], Rangachar Kasturi[1], Andrew Remsen[2], Kendra Daly[2]

sfefilatyev@gmail.com, kurtkramer@gmail.com , hall@cse.usf.edu, goldgof@cse.usf.edu, r1k@cse.usf.edu, aremsen@marine.usf.edu, kdaly@marine.usf.edu

[1]Department of Computer Science and Engineering,
University of South Florida,
Tampa, Florida, USA

[2]College of Marine Science,
University of South Florida,
St. Petersburg, Florida, USA

***Abstract*— The aim of this study is to investigate a data mining approach to help assess consequences of oil spills in the maritime environment. The approach under investigation is based on detecting suspected oil droplets in the water column adjacent to the Deepwater Horizon oil spill. Our method automatically detects particles in the water, classifies them and provides an interface for visual display. The particles can be plankton, marine snow, oil droplets and more. The focus of this approach is to generalize the methodology utilized for plankton classification using SIPPER (Shadow Imaging Particle Profiler and Evaluation Recorder). In this paper, we report on the application of image processing and machine learning techniques to discern suspected oil droplets from plankton and other particles present in the water. We train the classifier on the data obtained during one of the first research cruises to the site of the Deepwater Horizon oil spill. Suspected oil droplets were visually identified in SIPPER images by an expert. The classification accuracy of the suspected oil droplets is reported and analyzed. Our approach reliably finds oil when it is present. It also classifies some particles (air bubbles and some marine snow), up to 3.3%, as oil in clear water. You can reliably find oil by visually looking at the examples put in the oil class ordered by probability, in which case oil is found in the first 10% of images examined.**

Keywords - oil-droplet detection; images; classification; oil-spill; machine learning; support vector machine; plankton.

I. Introduction

The Deepwater Horizon Oil Spill is the biggest environmental disaster in the United States history and is the largest marine oil spill in the history of the oil industry [1]. The impact of the spill is still being evaluated with various estimates of the immediate damage, area affected, and longevity of its effect being generated [2]. Most of the studies are focused on oil that occurred on the ocean surface, addressing the fact that oil-mixtures often are lighter than water and, thus, tend to float on its surface. However, there are many indications that chemicals aimed to disperse the oil compounds turned the oil into neutrally buoyant oil droplets (see Figure 1a), which permeated the depths of the Gulf of Mexico [3]. The properties of such oil droplets have the potential to allow them particles to remain in the water for long periods of time, negatively affecting the marine habitat, fishing, and tourism industry.

In this study, we evaluate a special platform, SIPPER (Shadow Imaging Particle Profiler and Evaluation Recorder), for use in oil droplet detection in seawater. The SIPPER, which has been in use by marine scientists for the last decade, allows the timely extraction and identification of millions of plankton images per deployment as scanned by its underwater sensor. Based on a proven record for plankton population classification, we undertook a study to assess how suitable this platform is for detection of oil droplets suspected to be in the water.

Researchers from the University of South Florida's (USF) College of Marine Science collected image data during one of the first research cruises to the area affected by the Deepwater Horizon oil spill. The data include images of particles and plankton (see Figure 1b), along with suspected oil droplets. It is stressed here, that the oil droplets are only "suspected" as we have only images, but not corresponding physical water samples. However, based on the extensive experience of the marine scientists involved in the manual examinations of the data, it is believed that it is highly likely that the images represent oil droplets.

The aim of this research is to evaluate the effectiveness of image processing and machine learning techniques to process a large quantity of data and to classify particles obtained from the underwater research instrument, assuming the image data collected during the initial deployment indeed includes oil droplets. We are not aiming to draw any conclusions on the ecological meaning of the SIPPER data and presence of actual oil. However, this research may result in a vision-based method to assess the presence of such oil droplets in the water columns using the SIPPER imaging system. This paper briefly describes the hardware of the instrument, the algorithmic process used to discern the suspected oil droplets from other plankton particles, and the results obtained on a dataset collected in the immediate vicinity of the oil spill, as well as the results obtained from data collected from unaffected areas. We discuss our observations, limitations of the approach, and provide suggestions for further research.

978-0-7695-4409-0/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDMW.2011.65

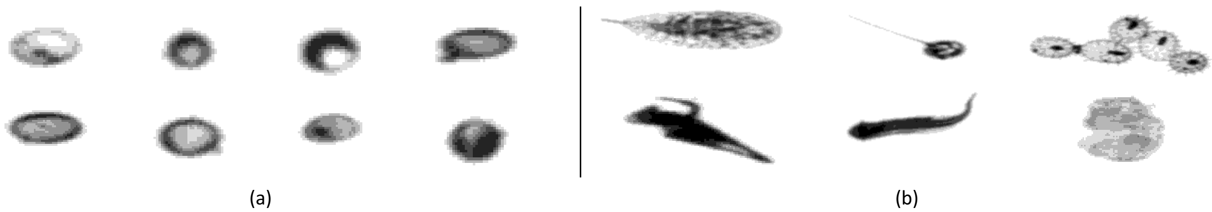

Figure 1. In-flow particles as imaged by SIPPER. (a) Suspected oil droplets found in the area of Deepwater Horizon oil spill. (b) Images of various particle classes from regular plankton population. Scale of each particle was changed to fit in the table.

II. DESCRIPTION OF PLATFORM AND ALGOIRTHMS

The SIPPER [4] was developed by the Center of Ocean Technology at the University of South Florida for the purpose of monitoring the composition, distribution and size structure of plankton and other suspended particles in aquatic environments. The SIPPER uses collimated LED illumination and a high speed line scan camera to continuously image particles and plankton as they pass through a 10cm × 10cm sampling aperture [5, 6]. A continuously scanning line scan camera captures images that are 10 cm in width and continuous in length. All resolvable particles that enter the sampling tube are imaged and saved as a single large SIPPER file with concurrently collected environmental data, such as temperature and depth embedded within the SIPPER file. A single six hour deployment can result in hundreds of thousands to millions of individual extracted particle images larger than 0.4 mm equivalent spherical diameter (ESD).

Custom designed software, the Plankton Imaging Classification Extraction System (PICES), was developed to quickly extract, classify, manage and analyze these discrete plankton images. A database management subsystem within PICES allows management of the large amount of data generated by SIPPER. PICES provides quick retrieval and organization of data by multiple parameters, such as, cruise, deployment, depth, salinity, temperature, class, date-time, etc.

Use of PICES results in efficient and timely processing of collected data. The algorithms used during processing of the data include those for image extraction, feature calculation, and image classification.

PICES uses a simple algorithm to extract images of separate particles based on foreground-background segmentation and a connected components algorithm. After segmenting the image of a particle, a number of features are calculated/extracted and a feature vector is created. The features are used by a classification algorithm in order to assign a class label to the image.

PICES uses a trained SVM [7] to classify particles in a supervised manner. The SVM classifier was chosen for several reasons. First, experiments with different classifiers on data from the SIPPER device showed that an SVM was more accurate than other classifiers [8]. Second, the SVM classifier provides a confidence or probability value [9] for its selection, giving more flexibility in the final decision process. Third, the SVM classifier is characterized by a small number of 'support vectors', the instances of the training set that lie on the border between classes. By reviewing those instances it is possible to detect (and later remove) incorrectly labeled training data, which can greatly affect the resultant classifier.

The training dataset is created from data labeled by an expert. One or more marine scientists view the images from one or more cruises and/or deployments and label some proportion of the images. One of the inherent disadvantages of an SVM is that it can only handle two-class problems. However, it can be extended to a multiple-class problem by using several strategies, for example one-versus-all or one-versus-one. Here, we used a one-versus-one strategy for every possible two-class combination by selecting the winning class using voting. Features for each class pairing were selected separately using a Binary Feature Selection (BFS) process described in [10].

III. DESCRIPTION OF FEATURES

During its operation, SIPPER records environmental data, such as water temperature, depth at which SIPPER was operating, oxygen concentration, salinity, and florescence. However, it was decided to use only image-related features as environmental data did not provide enough variety in our dataset. Some of the reasons for that are the following:

- The depth sensor is always used. However, the oil dataset was collected only for certain depth intervals. Hence, this may spuriously be used to indicate the presence of oil.
- Salinity, oxygen, and florescence sensors were covered during the collection of the oil dataset because of the fear of being contaminated by oil. Thus in the dataset collected in the vicinity of the oil spill, the values for these features do not exist.
- The temperature feature is not reliable because of the fact that the oil was only detected during one of the trips spanning a very limited time in relation to all possible temperature ranges in that region during the year. Thus, temperature might be chosen by the classifier to represent oil spuriously. Oil only occurred at a couple of temperatures, but could occur at any. Generally, this feature is useful because some plankton may prefer specific temperatures during different seasons.

Table I shows the image-related features that were used during the feature selection process. In total 93 features were used for describing the data. They included 82 features, which were previously designed for identifying the general plankton population. These features were mostly concerned with direct measurements (pixel count, intensity), geometric, and boundary properties. Another 11 features were

TABLE I. IMAGE FEATURES USED TO CLASSIFY PARTICLES PRESENT IN THE WATER

Category	Sub-Category	Feature Count
Moment Features [11]	Binary	8
	Intensity weighted	8
	Edge pixels only	8
Morphological		9
Head/Tail	Pixel counts of first quarter and last quarter	2
	Length vs. width	1
	Length	1
	Width	1
Filled Area		1
Convex Area		1
Transparency	Binary/Weighted	2
Texture Using Fourier Transform [12]	According to each frequency range	5
Contour Fourier	Average of Five Frequency Domains	5
	Hybrid combinations	15
Intensity Histogram	Without white space	7
	With white space	8
Circularity	Circularity, Equivalent Diameter, Eccentricity, ratios, etc	5
Texture	Intensity statistics, Smoothness, Uniformity, Entropy	6

specifically designed to aid in detection of oil droplets. They were mostly concerned with the circularity of the shape of oil droplets and their texture properties.

IV. SVM PARAMETER TUNING AND FEATURE SELECTION

In this work we used a one-versus-one strategy in order to implement a multi-class classifier. The primary reason for such a decision was the faster speed of training, which is shown by some studies [13]. In this strategy, all SVM classifiers for each possible binary class combinations were created. A class label is selected by a majority vote. In the case of a tie among classes, the probability parameter of SVM was used to select the class label.

The feature selection process consisted of two steps: initial SVM parameter tuning and binary feature selection. The parameters (gamma, C, A) of the SVM were optimized by performing a grid-search with a certain interval across the training dataset [14]. Using the SVM parameters determined in the first stage of the selection process, binary class feature selection was performed using wrappers [15, 16]. Each combination of features and SVM parameters was evaluated using 5-fold cross validation [17] and the classification accuracy on the training set was used to guide the selection process further. In cases, where the classification accuracy is equal for several evaluated sets, the correctness of probability (CPP) [9] was used to rank the sets.

V. DESCRIPTION OF DATA

The data consists of images collected by the SIPPER during one of the first research cruises to the area of the Deepwater Horizon (DWH) oil spill on May 5-16, 2010 on the USF research vessel *Weatherbird II*. Data from three deployments was collected within 5 km of the original site of the DWH platform on May 14 and 15. However, the ship was not allowed to closely approach the spill source per the Coast Guard's interpretation of safety in the region. Anomalous semi-opaque spherical particles were manually detected in the SIPPER images in the upper 10 m during these three deployments. These particles were imaged in areas where oil was visibly observed at the sea surface, during a time of relatively strong winds and building seas. These conditions provide a possible mechanism by which surface oil could be mixed down into the water column. Based on these results and because these particles were not observed in imagery collected in nearby waters where surface oil was not present, we labeled these particles as suspected oil droplets. They did not resemble other spherical particles that had previously been imaged by SIPPER, such as fish eggs, sarcodine protists, or air bubbles. Other cruises around this time resulted in the collection of SIPPER plankton data from areas of the Gulf unaffected by the oil spill. The data from unaffected areas was used to assess the sensitivity of the approach to the presence of oil droplets and compare the distributions of particles between the areas. Results of such comparisons may be used for future studies of the ecological impact of the spill.

Evaluation of the observed image data suggests that the water column contained mostly small particles. Many smaller particles are found for each large particle encountered. Figure 2 shows the size distribution of particles found in the SIPPER images from the research cruise to the by the oil spill area. Size is the area in pixels of each particle. For this study, the dimensions of each pixel are approximately 27 µm on each side.

Particles that exceeded 100 pixels in total area were extracted. As seen in Figure 2, there were abundant small particles present, while larger particles were far less numerous. However, due to the lack of resolvable features for the smallest extracted particle images, only particles greater than 250 pixels in total area were classified by an expert. It was decided to disregard all images smaller than 250 pixels to increase the accuracy of particle classification. Images of sizes > 250 pixels, according to our observations, contained enough texture and contour information to effectively differentiate among classes.

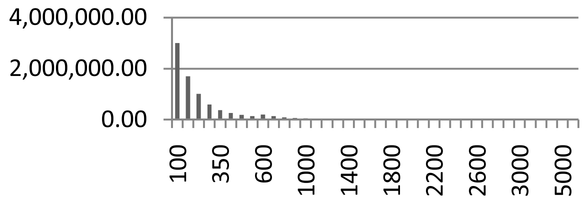

Figure 2. Size distribution of particles in the flow of water according to their size in pixels during the research cruise to the area of DWH oil spill.

TABLE III. DATASETS USED IN EXPERIMENTS

Dataset Identifier	Particles total	Oil droplets
Oil Original Set	8536	1072
Oil Original Replaced Set	8536	1072
Oil Large Test Set	43816	13858
Non-Oil Set	6745	0
Oil Random Set	13678	79

For all datasets, the number of classes was set to 36, which represented only major classes of particles with at least 20 instances. Table II shows the categories of classes used in our study. We created five datasets (see Table III) to study the data from the area affected by the oil spill, as well as unaffected areas.

The set of images, called *Oil Original Set*, was obtained by selecting instances of particles from the data obtained in the area affected by the spill, that were of primary interest to marine scientists. There were a number of selection criteria. First, the particles had to be identifiable in the sense that they had a high probability of being a particular plankton class or oil droplet. Second, since we were primarily interested in oil detection, oil droplets were a focus of the initial search and more likely to be labeled. This labeling was done first after the cruise. Thus, this dataset does not represent a completely random choice of particles. Overall, the set is composed of 8,537 particles, which represented less than 0.5% of all data during the cruise to the affected area. The oil droplet class was 1,072 instances, comprising 12.49% of the particles in the dataset. The decision to label each particle was made based on a visual analysis of the particle, with the knowledge available to the expert.

The *Non Oil Set* contains data that was collected during other cruises to areas unaffected by the BP oil spill. This data was collected from several locations in the Gulf of Mexico as well as the Caribbean Sea in 2010. The experiments conducted on this data were designed to test the classifier's specificity to oil droplets, i.e., detection of oil when no oil is present. The dataset had 6,745 particles belonging to 36 classes with zero oil droplets.

The *Oil Random Set* was created in a different manner. Instances of all particles in the set were randomly selected from the data from the cruise to the areas affected by the spill, not just by selecting particles of interest. The particles in the set were assigned the appropriate class based on visual analysis. Thus, this dataset has approximately the same distribution as the real distribution of particles in the water during that deployment.

TABLE II. CATEGORIES OF CLASSES OF PARTICLES USED IN EXPERIMENTS

Category of classes of particles	# of classes in category
Crustacean Copepod	5
Crustacean Eumalacostracan	3
Detritus (including oil droplet)	5
Echinoderm	1
Elongate	2
Fish	1
Gelatinous	8
Mollusc	2
Noise	2
Phytoplankton	1
Protist	5
Radiolarian	1
TOTAL	36

The next two datasets actually represent two categories of datasets, such that each dataset in a single category differs from other datasets in the same category only by the composition of its instances of oil droplet class. The datasets were obtained from the data from the same cruise in the following manner. First, the *Oil Original Set* was used to train an SVM classifier within PICES. The resulting classifier was used to classify all the data from the cruise. About 50,000 images that had a high confidence for the predicted class were viewed by an expert and given final class labels (which could be the same as the predicted class label). Some of the instances in this classified and validated data were part of the *Oil Original Set*, because it came from the same pool of raw data. Since our interest was mainly in oil detection, priority was given to the validation of oil droplet predictions. Out of 50,000 instances of classified and validated data about 20,000 instances were oil droplets.

The *Oil Original Replaced Set* is a category of datasets, which was obtained from the *Oil Original Set* by replacing the 1,072 oil droplets with oil droplets randomly selected from the set of classified and validated data as described above. The category has 30 datasets. Each of the 30 datasets in the category has the same number of instances as the *Oil Original Set*, but a different random set of instances from the oil droplet class. The experimental results are reported for the whole category *Oil Original Set Replaced* by averaging the results from each dataset within the category. For example, when it is stated that the classifier was trained on the *Oil Original Set Replaced* and tested on the *Non Oil Set* it means that 30 classifiers have been created. Each classifier was created by training on a unique dataset from the *Oil Original Set Replaced* category. Then, each classifier was used to classify the *Non Oil Set*. The results from 30 classifications are averaged and reported as a single experiment. By averaging the results from a statistically significant number of datasets, it is possible to minimize the risk of a good/bad random selection of oil droplet instances which can skew performance.

TABLE IV. PERFORMANCE OF SINGLE-STAGE CLASSIFIER. 10-FOLD CROSS VALIDATION ON OIL ORIGINAL SET

Oil droplet detection accuracy: 90.95%			
Absolute Performance:			
	Oil droplet	Other	Count:
Oil droplet:	975	97	1072
Other	104	7360	7464
Total:	1079	7457	8536
Relative Performance:			
	Oil droplet	Other	
Oil droplet:	90.95%	9.05%	
Other	1.39%	98.61%	

TABLE V. PERFORMANCE OF SINGLE-STAGE CLASSIFIER. 10-FOLD CROSS VALIDATION ON 30 SETS OF OIL ORIGINAL REPLACED SET CATEGORY.

Oil droplet detection accuracy: 94.37%			
Absolute performance:			
	Oil droplet	Other	Count:
Oil droplet	1011	61	1072
Other	94	7370	7464
Total:	1104	7431	8536
Relative Performance			
	Oil droplet	Other	
Oil droplet	94.37%	5.63%	
Other	1.25%	98.75%	

TABLE VI. PERFORMANCE OF CLASSIFIER. TESTED ON OIL LARGE TEST SET, TRAINED ON OIL ORIGINAL SET.

Oil droplet detection accuracy: 92.67%			
Absolute performance:			
	Oil droplet	Other	Count
Oil droplet	12842	1016	13858
other	1324	28634	29958
Total	14166	29650	43816
Relative Performance:			
	Oil droplet	Other	
Oil droplet	92.67%	7.33%	
Other	4.42%	95.58%	

TABLE VII. AVERAGE PERFORMANCE OF CLASSIFIER. TESTED ON OIL LARGE TEST SET, TRAINED ON OIL ORIGINAL REPLACED SET.

Oil droplet detection accuracy: 94.28%			
Absolute performance:			
	Oil droplet	Other	Count
Oil droplet	13066	792	13858
other	1013	28945	29958
Total	14079	29737	43816
Relative Performance:			
	Oil droplet	Other	
Oil droplet	94.28%	5.72%	
Other	3.38%	96.62%	

Similarly, the *Oil Large Test Set* is a category of 30 datasets. Each dataset in the category was obtained by including all predicted and validated data that passed through the following filters. Instances of the data that are a part of the *Oil Original Set* were removed. 5,000 images of oil droplets, selected randomly, were removed for future use for validation. Another 1,072 oil droplets used for building a particular dataset, within the category of the *Oil Original Replaced Set*, were removed as well. They are the 1,072 used to build the classifier being tested on this data set. Each dataset within the *Oil Large Test Set* category had 36 classes, 43,816 images total, of which 13,858 were oil droplets. The results of experiments where the category *Oil Large Test Set* was used are reported by averaging the performance from 30 individual classifiers within the category. For a particular experiment the classifier is trained on the *Oil Original Set Replaced* and tested on the *Oil Large Test Set,* such that instances of the oil droplet class in training and test datasets don't intersect.

The datasets *Oil Set Original*, *Oil Large Test Set,* and *Oil Random Set* did not intersect. Datasets *Oil Set Original* and *Oil Set Original Replaced* intersected for instances of all classes except oil droplets.

VI. EXPERIMENTS

In our experiments, we report the accuracy of classification in the form of a 2x2 confusion matrix, as if we were doing binary classification, although the setup of the experiment itself was not binary. One class was the oil droplet class, particles of particular interest for this research. The category 'other' represents the classification of all other particles compared against oil droplets. Thus, every prediction in favor of one of the other 35 classes of the datasets is summarized into the 'other' category. We do not report the accuracy among the 35 non-oil classes.

Binary Feature Selection was done to select features for each of the 630 binary SVM classifiers that comprised our one-stage classifier for 36 classes. Table IV shows the performance of the classifier using 10-fold cross validation on the *Oil Original Set*. The oil identification accuracy was 90%, with a less than 2% false positive rate. Table V shows the results of a 10-fold cross-validation on the *Oil Original Replaced Set*. The accuracy and false positive rate in that experiment were improved, correctly identifying 94.37% of oil droplets, with a 1.25% false positive rate. Such performance can be explained by the fact that the oil droplets in the *Oil Original Replaced Set* are likely more varied in their appearance. The original set of oil droplets included those that were most clear to the expert. Hence, we can expect that they appear quite homogeneous with minimal variation, so they could be sure of the label.

In all other experiments we report the performance of two classifiers, first trained on the *Oil Original Set* (called *Classifier I*) and second trained on the *Oil Original Replaced Set* (called *Classifier II*) to compare their sensitivity and specificity. As stated in the previous section, experiments with the *Oil Original Replaced Set* (*Classifier II*) involved 30 datasets that belonged to that category and the results

TABLE VIII. PERFORMANCE OF CLASSIFIER. TESTING ON NON OIL SET WHEN TRAINED ON OIL ORIGINAL SET.

False positive rate: 7.62%			
Absolute performance:			
	Oil droplet	Other	Count:
Oil droplet	0	0	0
Other	514	6231	6745
Total:	514	6231	6745
Relative Performance :			
	Oil droplet	Other	
Oil droplet	0.00%	0.00%	
Other	7.62%	92.38%	

TABLE IX. AVERAGE PERFORMANCE OF CLASSIFIER. TESTING ON NON OIL SET WHEN TRAINED ON OIL ORIGINAL REPLACED SET.

False positive rate: 3.29%			
Absolute performance:			
	Oil droplet	Other	Count:
Oil droplet	0	0	0
Other	222	6524	6745
Total:	222	6524	6745
Relative Performance :			
	Oil droplet	Other	
Oil droplet	0.00%	0.00%	
Other	3.29%	96.71%	

TABLE X. PERFORMANCE OF CLASSIFIER. TESTED ON OIL RANDOM SET, TRAINED ON OIL ORIGINAL SET.

Oil droplet detection accuracy: 75.95%			
Absolute performance:			
	Oil droplet	Other	Count
Oil droplet	60	19	79
Other	691	12908	13599
Total	751	12927	13678
Relative Performance :			
	Oil droplet	Other	
Oil droplet	75.95%	24.05%	
Other	5.08%	94.92%	

TABLE XI. AVERAGE PERFORMANCE OF CLASSIFIER. TESTED ON OIL RANDOM SET, TRAINED ON OIL ORIGINAL REPLACED SET.

Oil droplet detection accuracy: 63.92%			
Absolute performance:			
	Oil droplet	Other	Count
Oil droplet	51	28	79
Other	232	13367	13599
Total	283	13395	13678
Relative Performance :			
	Oil droplet	Other	
Oil droplet	63.92%	36.08%	
Other	1.71%	98.29%	

from those 30 runs were averaged. Thus, *Classifier II* actually represents 30 classifiers, but in the paper, for simplicity, we refer to it as to a single classifier.

The two classifiers were created and then applied to make classifications on the *Oil Large Test Set*, *Non Oil Set*, and *Oil Random Set*. The results for these experiments are shown in Tables VI-XI.

For the largest of our test sets, *Oil Large Test Set*, *Classifier I* achieved a detection rate of 92.67%, and *Classifier II* achieved 94.28%. The false positive rates were 4.42% and 3.38%, respectively (see Tables VI and VII). Thus, *Classifier II* had better accuracy of detection and false positive rate. Similar performance in relation to the false positive rate was observed in experiments with the *Non Oil Set* (see Tables VIII and IX). *Classifier II* had a false positive rate of 3.29% as opposed to 7.62% using *Classifier I*. A greater performance difference between *Classifier I* and *Classifier II* was observed while testing on the *Oil Random Set*. The detection rate for oil droplets with *Classifier I* was 75.95% and the false positive rate was 5.08% (see Table X). The detection rate with *Classifier II* was 12% lower in this case, 63.92%, with about one third as many false positives, 1.71% (see Table XI). Overall, the detection rate was lower for the *Oil Random Set*, than with all previous test datasets.

Because of the way the *Oil Random Set* was built, it had a distribution of particles similar to that expected in the vicinity of the oil spill. In examining the current SIPPER data, one finds the percentage of oil droplets in the dataset was about 0.5%. The false positive rate for both classifiers was always higher (1.25-7%). Thus, for regular SIPPER data, it is not yet possible to automatically verify the presence of oil droplets in water with the currently built classifiers.

Consequently, we took the class predictions from *Classifier II* on the *Oil Random Set* and extracted probabilities for them from a modified version of *libsvm* [10]. We then ranked the examples classified as oil by probability from highest to lowest (Figure 3). These results indicate that the number of oil droplets is always between 11 and 25% of the predicted oil. The good news is that if an expert looks at the images classified as oil, they will find some oil in the top 10% and top 20% of the classifications (see Figure 4). If they were to randomly search through images with 0.5% oil, when oil droplets are present, they would need to look at 200 examples to find one oil sample. They will find four in the first 25 examined with our tool. Thus, the user can quit looking if no oil is found in the first 50 or so images that are highly ranked by probability of being oil.

Now, it is clear from looking at Figures 3 and 4 that many of the top probability “oil droplets” are, in fact, not oil. Air bubbles and marine snow can look very similar. In Figure 4, the non-oil images are a little more elliptical in shape than oil droplets. However, oil does not have to be perfectly spherical as we can see from Figure 1. Reviewing the features selected for each individual binary SVM classifier that comprised this single-stage classifier, it was confirmed that the most important features used to

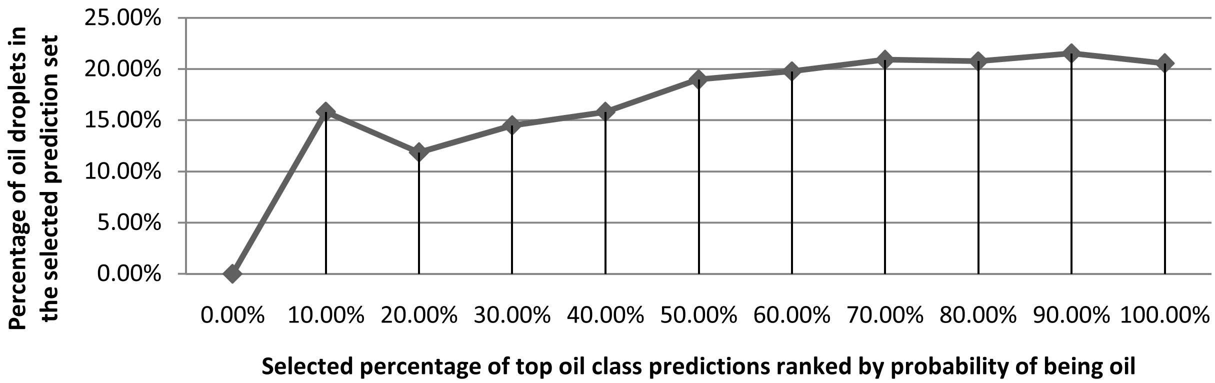

Figure 3. Percentage of oil droplets in the predictions when sorted by probability of being oil. So, 10% means the 25 highest probability predictions for oil of which 4 are actually oil. This is with Classifier II applied to the oil random dataset with results shown in Table 11.

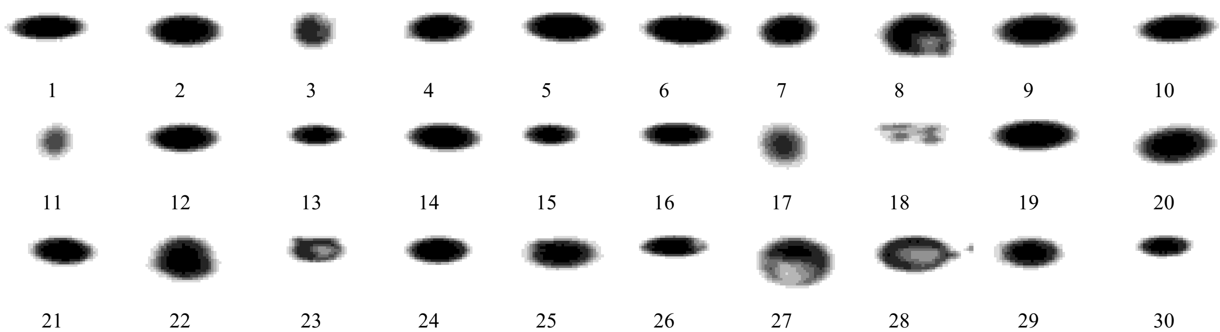

Figure 4. Top 30 particles classified as oil droplets when sorted by the probability. Particles numbered 3, 8, 23, 27, 28 are suspected oil droplets. Particle number 18 is detritus snow. The others are noise bubbles

discriminate the oil droplet class from others were related to the circularity of the shape, and the texture of the particle. However, it turned out it is not quite enough for completely automatic oil detection. The current performance can be further explained by the following. Reviewing the support vectors for the binary classifier, which generates the most false positives – Oil Droplet vs. Noise Bubbles we found incorrectly labeled training data, which was used to characterize the classifier. In the list of 18 support vectors that constituted the border of the classifier that separated the oil droplets, four of the support vectors were in fact noise bubbles instead of oil droplets. This fact was confirmed by an expert when instances of training data of the oil droplet class, constituting the support vectors for the mentioned classifier, were re-examined. Further improvements are likely possible from having an expert examine the support vectors of the most confused classes.

VII. Discussion

The analysis of the particles, which were most confused with oil droplets, suggests that there are only three major classes that have an appearance similar to oil droplets: detritus snow, noise bubbles, and protists with a lopsided shape. It is possible that a two-stage classifier will allow fully automatic detection of oil droplets in water near the spill. The first stage of the classifier would be aimed at sensitivity to oil droplets, while producing many false positives. The second stage of the classifier would focus on specificity to oil droplets. This setup is reported to be useful to detect very rare events and for cases in which many features are costly to compute [18]. Certain improvements can be also made in relation to features used to discriminate between the most confused classes. Circularity features were found very useful to discriminate oil droplets, which are often circular in shape, from many plankton organisms. However, those features are not particularly useful with other classes showing circularity – noise bubbles and marine snow. Further, we have found that our experts use depth to help them classify oil. We did not use this feature, as many of the non-image features were not available due to precautions made to protect the environmental sensors from oil damage.

VIII. Conclusions

Overall, a trained SVM achieved a high detection rate for oil droplets. When tested on the *Oil Large Test Set*, consisting of 43,816 particles, of which 13,858 were oil droplets, the accuracy of detection was over 94%, which is comparable to the cross validation test on the training set.

The false positive rate was less than 3.4% in all experiments with *Classifier II*, which was trained on a random selection of oil examples. We did an experiment with a randomly chosen test set, whose distribution mimicked what would be expected during the cruise (about 0.5% oil). For that dataset, oil droplet detection was just 63%. It is also the case that in water where there was no oil, our classifiers predicted that a small amount of oil was present.

We showed that by using probabilities for the class predictions and ordering them from highest to lowest, oil will regularly appear in the top 10-30% of data. So, if an expert uses our tool, PICES, to view the images that are predicted to be oil, they will be able to reliably find oil droplets more quickly than randomly searching through particles.

There is room for improved oil detection to enable the best analysis of how much oil is in the water. This will occur through classifier tuning/replacement and new features. It is also important to review the particular instances of training data that are selected as support vectors for binary classifiers. The current results are promising in terms of observing subsurface oil and obtaining a general count of the number of oil droplets.

Acknowledgment

This research was made possible by a grant from BP/The Gulf of Mexico Research Initiative through the Florida Institute of Oceanography and Office of Naval Research grant N00014-07-0802.

References

[1] Deepwater Horizon Unified Command, US Scientific Teams Refine Estimates of Oil Flow from BP's Well Prior o Capping, *Gulf of Mexico Oil Spill Response 2010*; available from: www.deepwaterhorizonresponse.com/go/doc/2931/840475

[2] Camilli, R. Reddy, C.M. Yoerger, D.R. Van Mooy, B.A.S. Jakuba, M.V., Kinsey, J.C. McIntyre, C.P., Sylva, S.P. and Maloney, J.V., “Tracking hydrocarbon plume transport and biodegradation at Deepwater Horizon”, *Science*, v.330, n.6001, pp.201, 2010

[3] Dykes, B.M.,”Researchers find thick patches of crude still on Gulf floor”, *Yahoo News*. Retrieved on 2011-04-26

[4] Samson S., Hopkins, T., Remsen, A., Langebrake, L., Sutton, T., and Patten, J., “A System for High-Resolution Zooplankton Imaging”, *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, October 2001

[5] Remsen, A., Hopkins, T., and Samson, S., “What You See is Not What You Catch: A Comparison of Concurrently Collected Net, Optical Plankton Counter, and Shadowed Image Particle Profiling Evaluation Recorder Data from Northeast Gulf of Mexico”, *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 51, no. 1, pp. 129-151, 2004

[6] Remsen, A., “Evolution and field application of a plankton imaging system”, *Ph.D. Dissertation*, College of Marine Science, University of South Florida, 2008

[7] Burges, C., “A Tutorial on Support Vector Machines for Pattern Recognition”, *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, June 1998

[8] Luo T., Kramer, K., Goldgof, D., Hall, L., Samson, S., Remsen, A., Hopkins, T., “Recognizing Plankton Images From the Shadow Image Particle Profiling Evaluation Recorder”, *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, vol..34, no. 4, pp. 1753-1762.

[9] Platt J., “Probabilistic Outputs for Support Vector Machines and Comparison to Regularized Likelihood Methods” *in Advances in Large Margin Classifiers*, pp. 61-74, Cambridge, MA, USA, 1999

[10] Kurt, K., “System for Identifying Plankton from the SIPPER Instrument Platform”, *Doctoral Dissertation*, University of South Florida, 2010.

[11] Hu, M.K., “Visual Pattern Recognition by Moment Invariants”, *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179-187, 1962

[12] Zhang, D. and Lu, G., “A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures*”, Journal of Visual Communications and Image Representation*, vol. 14, no. 1, pp. 41-60, 2003

[13] Hsu, C. and Lin C., “A Comparison of Methods for Multi-Class Support Vector Machines”, *IEEE Transactions on Neural Networks*, vol. 13, no.2, pp. 415-425, March 2002

[14] Staelin, C., “Parameter Selection for Support Vector Machines”, *HP Laboratories Israel*, Technion City, Haifa, 2002

[15] Kohavi, J.R. and George, H., “Wrappers for Feature Subset Selection”, *Artificial Intelligence*, vol. 97, no. 1., pp. 273-324, December 1997

[16] Silva, H. and Fred, A., “Pair wise vs. Global Multi-Class Wrapper Feature Selection” in *Proceedings of the 6th Conference on 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases* (AIKED’07), vol. 6, Corfu Island, Greece, 2007, pp. 1-6

[17] Witten, I. and Frank, E., “Data Mining: Practical Machine Learning Tools and Techniques”, *Morgan Kaufmann Publishers*, 2005

[18] Senator, T.E., “Multi-stage classification”, *Proceedings of the Fifth IEEE International Conference on Data Mining*, 2005