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]

II. DESCRIPTION OF PLATFORM AND

The SIPPER [4] was developed by the Technology at the University of South purpose of monitoring the composition, dist structure of plankton and other suspended pa environments. The SIPPER uses c illumination and a high speed line s continuously image particles and plankto through a 10cm × 10cm sampling ape continuously scanning line scan camera cap are 10 cm in width and continuous in lengt particles that enter the sampling tube are in as a single large SIPPER file with concu environmental data, such as temperat embedded within the SIPPER file. A deployment can result in hundreds of thous of individual extracted particle images larg equivalent spherical diameter (ESD).

Custom designed software, the Pl Classification Extraction System (PICES), v quickly extract, classify, manage and analy plankton images. A database management s PICES allows management of the large generated by SIPPER. PICES provides qu organization of data by multiple parameters deployment, depth, salinity, temperature, etc.

Use of PICES results in efficient and t of collected data. The algorithms used duri the data include those for image ex calculation, and image classification.

PICES uses a simple algorithm to ex separate particles based on foregro segmentation and a connected components segmenting the image of a particle, a numb calculated/extracted and a feature vector features are used by a classification algor assign a class label to the image.

PICES uses a trained SVM [7] to class supervised manner. The SVM classifier several reasons. First, experiments with dif on data from the SIPPER device showed t more accurate than other classifiers [8]. S classifier provides a confidence or probabi

For all datasets, the number of classes was set to 36, 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*

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

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.