Recognizing Plankton Images From the Shadow Image Particle Profiling Evaluation Recorder

Tong Luo, Kurt Kramer, *Member, IEEE*, Dmitry B. Goldgof, *Senior Member, IEEE*, Lawrence O. Hall, *Fellow, IEEE*, Scott Samson, Andrew Remsen, and Thomas Hopkins

***Abstract*—We present a system to recognize underwater plankton images from the shadow image particle profiling evaluation recorder (SIPPER). The challenge of the SIPPER image set is that many images do not have clear contours. To address that, shape features that do not heavily depend on contour information were developed. A soft margin support vector machine (SVM) was used as the classifier. We developed a way to assign probability after multiclass SVM classification. Our approach achieved approximately 90% accuracy on a collection of plankton images. On another larger image set containing manually unidentifiable particles, it also provided 75.6% overall accuracy. The proposed approach was statistically significantly more accurate on the two data sets than a C4.5 decision tree and a cascade correlation neural network. The single SVM significantly outperformed ensembles of decision trees created by bagging and random forests on the smaller data set and was slightly better on the other data set. The 15-feature subset produced by our feature selection approach provided slightly better accuracy than using all 29 features. Our probability model gave us a reasonable rejection curve on the larger data set.**

***Index Terms*—Feature selection, learning, plankton recognition, probabilistic output, support vector machine (SVM).**

I. Introduction

RECENTLY, the shadow image particle profiling evaluation recorder (SIPPER) was developed to continuously sample plankton and suspended particles in the ocean [23]. The SIPPER uses high-speed digital line-scan cameras to record images of plankton and other particles, thus, avoiding the extensive post-processing necessary with analog video particle images. The large sampling aperture of the sensor combined with its high imaging resolution (50 $\mu\text{m p/pixel}$), means that it is capable of collecting tens of thousands of plankton images an hour. This soon overwhelms a scientist attempting to manually classify the images into recognizable plankton groups. Therefore, an automated plankton recognition system is necessary to solve the problem or at the very least to help with the classification.

Tang [25] developed a plankton recognition system to classify plankton images from video cameras. The moment invariants and Fourier descriptor features from contour images were extracted. Also, granulometric features from the gray-level images were computed. Finally, a learning vector quantization neural network was used to classify examples. Tang [25] achieved 92% classification accuracy on a medium-size data set.

The project Automatic Diatom Identification and Classification (ADIAC) has been ongoing in Europe since 1998. Different feature sets and classifiers have been experimented with to recognize separate species of diatoms taken from photo-microscopes. Loke [16] and Ciobanu [5] studied some new contour features. Santos [24] extended the contour features to multiscale Gabor features together with texture features. Wilkinson [29] applied morphological operators to help extract both contour and texture information. Fischer [11] summarized these features and used ensembles of decision trees to classify the combined feature set. Greater than 90% overall accuracy was achieved on the diatom images.

However, images from previous work are of relatively good quality or at least with clear contours. Therefore, complicated contour features and texture information can be extracted easily. The SIPPER images, on the other hand, present several of the following difficulties:

1) Many SIPPER images do not have clear contours. Some are partially occluded. Therefore, we cannot primarily depend on contour information to recognize the plankton.
2) The SIPPER image gallery includes many unidentifiable particles as well as many different types of plankton.
3) The SIPPER images in our experiments are binary, thus lacking texture information.

Tang [26] proposed several new features for SIPPER images and applied multilevel dominant eigenvector methods to select a best feature subset. A Gaussian classifier was employed to recognize the image features and validate the feature selection methods on selected identifiable plankton.

Not depending heavily on contour information, Luo [17] developed several special features, and applied a support vector machine (SVM) [27] to classify the feature vectors. This paper is an expansion of the work reported in [17]. We will describe our feature selection strategy-the wrapper approach [15] with backward elimination and a new way to compute probabilistic outputs in a multiclass SVM on the reduced feature subset.

This paper is organized as follows. Section II introduces the binary SIPPER images used in the experiments. In Section III, we discuss the preprocessing of the images and the extraction of the features. Section IV describes the SVM and the probability assignment in a multiclass SVM. We applied wrappers

Manuscript received September 4, 2003; revised March 2, 2004. This work was supported in part by the United States Navy, Office of Naval Research, under Grant N00014-02-1-0266 and the National Science Foundation under Grant EIA-0130768. This paper was recommended by Associate Editor P. Willet.

T. Luo, K. Kramer, D. B. Goldgof, and L. O. Hall are with the Department of Computer Science and Engineering, University of South Florida, Tampa, FL 33620 USA (e-mail: tluo2@csee.usf.edu; kkramer@csee.usf.edu; goldgof@csee.usf.edu; hall@csee.usf.edu).

S. Samson, A. Remsen, and T. Hopkins are with the College of Marine Science, University of South Florida, St. Petersburg, FL 33701 USA (e-mail: samson@seas.marine.usf.edu; aremsen@seas.marine.usf.edu; thopkins@seas.marine.usf.edu).

Digital Object Identifier 10.1109/TSMCB.2004.830340

1083-4419/04$20.00 © 2004 IEEE

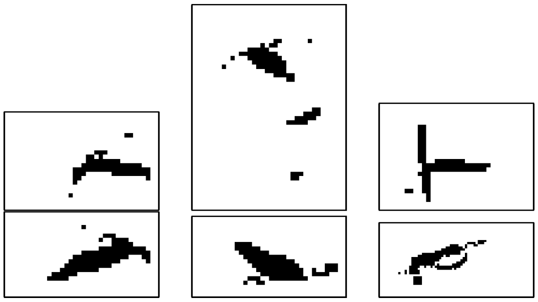


Fig. 1. Copepod.

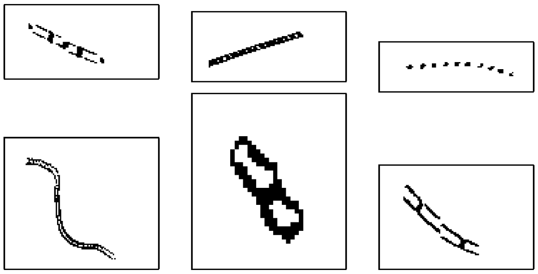


Fig. 2. Diatom.

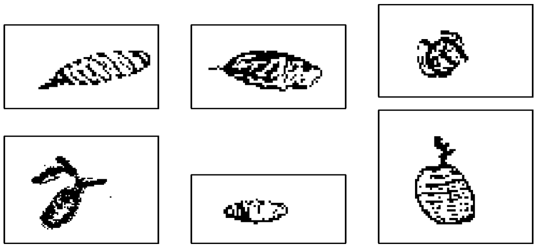


Fig. 3. Doliolid.

with backward elimination to select the best feature subset in Section V and experimental results for the system are detailed in Section VI. Finally we summarize our work and propose some ideas for future work in Section VII.

II. Image Gallery

The image gallery includes 7285 binary SIPPER images: 1285 images from five types of plankton were initially selected by marine scientists as our starting point. The other 6000 images were samples from a deployment of SIPPER from the Gulf of Mexico. The 6000 image were from the five most abundant types of plankton and manually unrecognizable particles. All the images were manually classified by marine scientists. Figs. 1–7 are typical examples of plankton and unidentifiable particles from the SIPPER image set.

III. Feature Computation

In the field of shape recognition, some general features like invariant moments, Fourier descriptors, and granulometric features etc. are widely used [7]. However, those general features are insufficient to capture the information contained in SIPPER images sampled from the Gulf of Mexico. Moreover, the SIPPER images have a lot of noise around or on the plankton and many images do not have clear contours, thus making contour features (Fourier descriptor [30], etc.) unstable and inaccurate. To solve this problem, we first preprocessed the images to suppress noise. We only extracted invariant moments and granulometric features, which are relatively stable with respect to noise and do not depend heavily on the contour image. To capture the specific information from our SIPPER image set, domain knowledge was used to extract some specific features such as size, convex ratio, transparency ratio, etc. There are 29 features in total as shown in Table I.

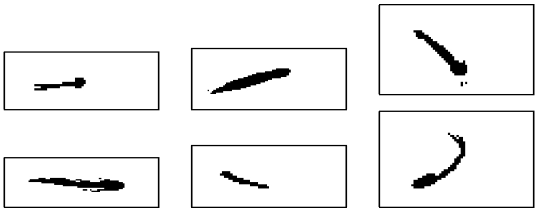


Fig. 4. Larvacean.

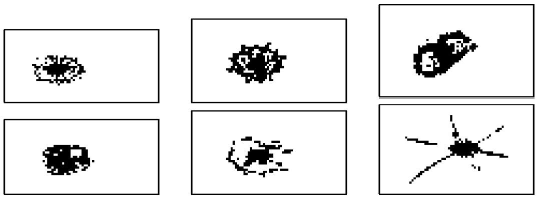


Fig. 5. Protoctista.

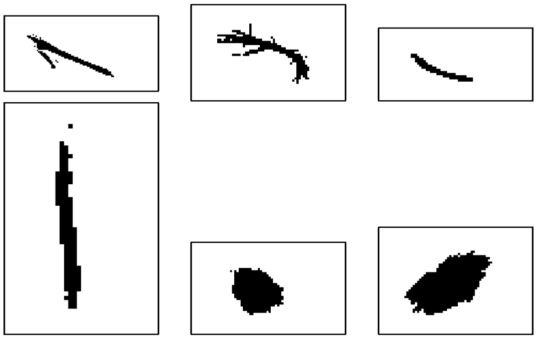


Fig. 6. Trichodesmium.

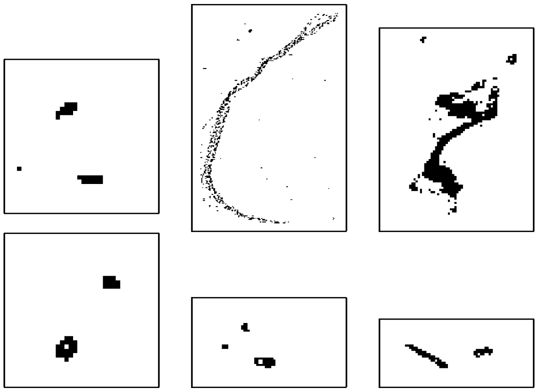


Fig. 7. Unidentifiable particles.

A. Object Detection and Noise Suppression

Marine scientists used a specialized software to detect objects. A series of morphological dilations were performed to

TABLE I
DESCRIPTION OF 29 FEATURES

Features	Number of features
Moment invariants of the original image	7
Moment invariants of the contour image after closing	7
Granulometric features	7
Domain specific features	8

connect nearest image pixels. If the bounding box of the connected image pixels after dilation was bigger than 15 × 15, the original image was stored as an object. Otherwise, the image pixels were considered irrelevant and deleted.

After detecting objects, we applied connected component analysis to eliminate the noise pixels far from the object bodies. Under the eight-connectivity condition (that is, all eight neighbor pixels of a pixel are considered connected to it), if a pixel's connected path to the image body is more than four, it will be regarded as noise and eliminated. In addition, a morphological closing operation with a 3 × 3 square window as the structure element was used to get a roughly smooth image shape and separate the holes inside the plankton body from the background [20]. This operation also helps to compute several domain specific features described in Section III-D and to get a rough contour of the image.

B. Moment Invariants

Moment features are widely used as general features in shape recognition. The standard central moments are computed as follows.

$(\bar{x}, \bar{y})$ is the center of the foreground pixels in the image. The $(p+q)$-order central moments are computed with every foreground pixel at (x, y)

$$\mu(p,q) = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q. \tag{1}$$

Then central moments are normalized by size as shown in (2)

$$\eta(p,q) = \frac{\mu(p,q)}{\mu(0,0)^{((p+q)/2+1)}}. \tag{2}$$

Hu [14] introduced a way to compute the seven lower order moment invariants based on several nonlinear combinations of the central moments. Using the normalized central moments, we got scale, rotation, and translation invariant features. We computed the same seven moment invariants on the whole object and the contour image after a morphological closing operation, respectively. See [14] for the detail description of the seven moments.

C. Granulometric Features

Since the Hu moments only contain low order information from the image, we also extracted granulometric features [18], which are robust measurements of the high order information. Granulometric features were computed by doing a series of morphological openings with different sizes of structure elements. Then we recorded the differences in size between the plankton before and after openings. Granulometric features are relatively robust to noise and contain the inherent information of shape distribution. Tang [25] found that granulometric features were the most important features in his experiments.

We applied 3×3, 5×5, 7×7, and 9×9 square windows as structure elements and did a series of morphological openings. Then differences in size were normalized by the original plankton size to obtain the granulometric features. Also, we applied 3×3, 5×5, and 7×7 square windows as structure elements, and did a series of morphological closings. The differences in size were normalized in the same way. We did not apply 9×9 square window to the closing because the SIPPER images are so small that most of them are diminished after the closing with a 7 × 7 square window as the structure element. The granulometric features are computed in (3)

$$g_i = \frac{\#\ \text{pixels changed after } i^{th} \text{ morphological operations}}{\#\ \text{pixels in the original image}}. \tag{3}$$

D. Domain Specific Features

Moment invariants and granulometries only capture some global information, which is insufficient to classify SIPPER images. Given advice from domain experts, we developed some domain specific features to help classification. The domain specific features include size, convex ratio, transparency ratio, eigenvalue ratio, and ratio between the plankton's head and tail.

- Size: It is the area of the plankton body, that is, the number of foreground pixels in the plankton image. The size features were extracted for both the original image and the contour images.
- Convex ratio: We implemented a fast algorithm [1] to get the convex hull of the plankton image. The convex ratio is the ratio between the plankton image size and the area of the convex hull. This feature contains information about the plankton boundary irregularity. We computed convex ratios with (4) for the original image and the image after a morphological opening. The morphological opening was to eliminate the noise around the plankton, which may cause an incorrect convex hull

$$cr = \frac{\#\ \text{pixels in the original image}}{\#\ \text{pixels in the convex hull}}. \tag{4}$$

- Transparency ratio: This is the ratio between the area of the plankton image and the area of the plankton after filling all inside holes. The transparency ratio helps in recognizing the transparent plankton. We computed the transparency ratios with (5) for both the original image and the image after a morphological opening. The morphological closing was used to remove the noise inside the plankton body

$$tr = \frac{\#\ \text{pixels in the original image}}{\#\ \text{pixels within the contour}}. \tag{5}$$

- Eigenvalue ratio: We first computed covariance matrix between the x and y coordinates of all pixels on the plankton bodies. Then the ratio between the two eigenvalues from the covariance matrix was calculated in (6). This ratio helps classify elongated plankton

$$er = \frac{\min(f_1, f_2)}{\max(f_1, f_2)} \tag{6}$$

where f_1, f_2 are eigenvalues of $\mathrm{cov(X, Y)}$.

- Ratio between the head and the tail: Some plankton such as larvaceans have a large head relative to their tail. We computed the ratio between the head and tail to differentiate them. To do this we first rotated the image to make the axis with the bigger eigenvalue parallel to the x-axis. Assuming the smallest and largest x values are 0 and T, respectively, we accumulated the number of foreground pixels along the x-axis from 0 to $(1/4)T$ and from $(3/4)T$ to T respectively. Then we computed the ratio between them as the ratio between the head and the tail.

IV. Support Vector Machines and Probability Model

SVMs [27] are receiving increasing attention these days and have achieved very good accuracy in pattern recognition, text classification, etc. [6]. In this section we describe SVMs and introduce a way to assign a probability value after multiclass SVM classification.

A. Support Vector Machines

SVMs, based on statistical learning theory, try to maximize the generalization accuracy using structure risk minimization [27]. The goal of classification is to minimize the generalization loss (normally the true misclassification rate). While most learning algorithms (decision trees, neural networks, etc.) try to minimize the empirical loss (normally the misclassification rate on the training set) with some criterion to keep reasonable generalization accuracy, SVMs minimize the empirical loss with a guaranteed upper bound of generalization loss.

Vapnik [27] introduced the Vapnik–Chervonenkis (VC) dimension to measure the classification capacity of a hypothesis space. Assuming that h is the VC dimension of the hypothesis space, m is the sample size, R_{true} is the true loss and R_{emp} is the empirical loss, the generalization bound holds with probability $1-\eta$ in (7)

$$R_{\text{true}} \le R_{\text{emp}} + \sqrt{\frac{h(\ln\frac{2m}{h}) + h - \ln\frac{\eta}{4}}{m}}. \tag{7}$$

In order to keep the generalization bound low, we have to minimize the empirical loss in a hypothesis space with a small VC dimension.

In binary classification, SVMs try to find a hyperplane to separate the data into two classes. In the case in which all the data are well separated, the margin is defined as two times the distance between the hyperplane and the closest example. SVMs search for the hyperplane with the largest margin, which has the lowest VC dimension. Therefore, SVMs ensure a low generalization loss bound. To increase the classification ability, SVMs first map the data into a higher dimension feature space with $\phi(x)$, then search for a hyperplane with the largest margin in that feature space to separate the data. In the feature mapping stage, the kernel $k(x,y) = \langle\phi(x)\cdot\phi(y)\rangle$ is used to avoid explicit inner product calculation in the high-dimension feature space.

C-SVM, a typical example of soft SVMs is described as follows.

Training Set: there are m examples: $x_1, x_2, \ldots, x_m$ with class label $y_i \in \{-1, 1\}$.

C-*SVM:*

$$\text{minimize } \frac{1}{2}\langle w, w\rangle + \frac{C}{m}\sum_{i=1}^{m}\xi_i \tag{8}$$

$$\text{subject to : } y_i(\langle w, \phi(x_i)\rangle + b) \ge 1 - \xi_i \tag{9}$$

$$C, \xi_i > 0 \tag{10}$$

where w is normal to the hyperplane, C is a scalar value that controls the trade off between the empirical risk and the margin length, ξ_i is the slack variable to handle nonseparable examples.

The decision function is $f(x) = \sum_i \alpha_i k(x_i, x) + b$, where α_i and b are computed from (8)–(10).

The Karush–Kuhn–Tucker condition of the optimal solution to (8) – (10)

$$\alpha_i[y_i(\langle w, \phi(x_i)\rangle + b) - 1 + \xi_i] = 0. \tag{11}$$

The variable α_i is nonzero only when (12) is satisfied. In this case x_i contributes to the decision function and is called a support vector (SV)

$$y_i(\langle w, \phi(x_i)\rangle + b) = 1 - \xi_i. \tag{12}$$

Therefore, we get a sparse solution of the decision function, where only SVs contribute.

There are two main approaches to extending SVMs to multiclass classification: one-versus-all and one-versus-one.

1) One-versus-all: A set of binary SVMs are trained to separate one class from the rest. The drawback is that we are handling unbalanced data when building binary SVMs. Moreover, each binary SVM is built on a totally different training set. There might be cases in which some binary SVMs conflict with each other for some examples. It is difficult to assign the class by just the real-valued outputs from every binary SVM.
2) One-versus-one: All possible groups of two classes are used to build binary SVMs. In the N class case, we will build $N(N-1)/2$ binary SVMs. When a new example is tested, all the binary SVMs vote to classify it.

The one-versus-one approch needs to build more binary SVMs than the one-versus-all approach, however, each of its binary SVMs only learns on a fraction of the data, thus can be time efficient in a large data set. Hsu [13] compared the one-versus-all and one-versus-one approach to handle multiple class problems in SVMs. They found the one-versus-one approach was much faster and more accurate than the one-versus-all approach on most data sets. We also compared the two approaches on our data sets in Sections VI-A and VI-B and the one-versus-one was superior to the other in our experiments.

B. Assigning Probability Values in Support Vector Machines

A probability associated with a classifier is often very useful and it gives some confidence about the classification result. For instance, the classifier could reject the example and leave it to a human to classify it when the confidence is very low. Platt [22] introduced the sigmoid function as the probability model to fit $P(y=1|f)$ directly. The parametric model is shown in (13)

$$P(y=1|f) = \frac{1}{1+\exp(Af+B)} \tag{13}$$

where A and B are scalar values. f is the decision function of a binary SVM.

A and B are fit with maximum likelihood estimation from the training set. Platt tested the model with three data sets including the UCI Adult data set and two other web classification data sets. The sigmoid-model SVM had good classification accuracy and probability quality in his experiments.

We followed the sigmoid model and extended it to the multiclass case. For practical reasons, the probability approximation needs to be fast since retraining for the probability model is frequently needed as more plankton images are acquired on a cruise. In the one-versus-one multiclass SVM model, since it is time consuming to do the parameter fitting for all $N(N-1)/2$ binary SVMs, we developed a practical approximation method to compute the probability value while avoiding parameter fitting.

1) We assume $P(y=1|f=0)=P(y=-1|f=0)=0.5$. This means that a point right on the decision boundary will have 0.5 probability of belonging to each class. We eliminate parameter B in this way.
2) Since each binary SVM has a different margin, a crucial criterion in assigning the probability, it is not fair to assign a probability without considering the margin. Therefore, the decision function $f(x)$ is normalized by its margin in each binary SVM. The probability model of SVMs is shown in **(14)**. P_{ij} represents the probability output for the binary SVM on class i versus class j, where class i is $+1$ and -1.

$$P_{ij}(y=1|f)=\frac{1}{1+\exp\left(\frac{-Af}{\|w\|}\right)} \tag{14}$$

$$P_{ij}(y=-1|f)=1-P_{ij}(y=1|f)=P_{ji}(y=1|f). \tag{15}$$

3) After we get the probability value for each binary SVM, the final probability for class i is computed as follows:

$$P(i)=\prod_{j}^{j\neq i}P_{ij}(y=1|f). \tag{16}$$

Normalize $P(i)$ to make $\sum_i P(i)=1$

4) output $k=\arg\max_i P(i)$ as the prediction.

Both Platt's approach for two-class problems (search for A and B) and our approach for multiple class problems (search for A) try to minimize log-likelihood loss function as in (17).

$$-\log L=-\sum_i t_i P(t_i) \text{ where } t_i \text{ is the class label.} \tag{17}$$

Since the loss function is not convex, we used line search for a single parameter A to avoid local minima. We also compared it with gradient descent search for A and B as Platt proposed. The comparison will be detailed in Section VI-D.

After learning a SVM model and setting a rejection threshold p, we reject an example and leave it to be classified by a person if $P(k)<p$.

V. Feature Selection

Feature selection helps reduce the feature computation time and increase the accuracy. There are two basic ways to do feature selection [8]. The filtering approach attempts to select a subset of features without applying learning algorithms. It is fast, but seems unlikely to result in the best accuracy. The wrapper approach [15] selects a feature subset by applying the learning algorithm. It has the potential to result in very good accuracy but is computationally expensive. A feature selection method specifically for SVMs was proposed recently. Weston [28] tried to minimize the generalization bound by minimizing the radius of the sphere including all the training examples. The drawback of this approach is that the generalization bound is loose, and minimizing the loose bound may not provide a feature subset with good accuracy.

In our system, we applied the wrapper approach with backward elimination. Backward elimination means one starts with all the features and systematically eliminates features. The average accuracy from a fivefold cross validation was used as an evaluation function. In our case, we start with 29 features and remove one feature from the feature set and get 29 different feature subsets with 28 features. We evaluate the 29 feature subsets by running fivefold cross validation and choose the feature subset with best average accuracy to explore. For instance, if the feature subset with best average accuracy is M, we remove one more feature from M and get 28 feature subsets with 27 features to add to the remaining candidate feature subsets. In this way, we can use certain search strategies to explore those feature subsets and keep removing features.

The algorithm halts if there is no improvement in accuracy for p successive feature subsets explored. Best first search (BFS), which is embedded in the wrapper approach, is used to explore the feature subset space, although it tends to stop with many features. In order to explore feature subsets with small numbers of features, greedy beam search (GBS) [12] was employed on the final feature subsets selected by BFS. GBS operates by only expanding the best q (beam width) leaf-nodes without any backtracking. It can quickly reduce the number of features to one.

To reduce the effect of overfitting, we took 20 percent of the data as a held-out data set, and did the feature selection on the remaining data while testing the selected feature subsets on the held-out data. The feature selection procedure is described in Algorithm 1.

```
Algorithm 1 Feature Selection Algorithm
1: N = $\{a_1, a_2, \ldots, a_{29}\}$, S = {N}, T = $\emptyset$, $q$ = constant, max = 0 where $a_i$ is the $i$th feature, S is a sorted list, max is the maximum accuracy, and $q$ is the beam width.
2: Compute the average accuracy $f(N)$ from a fivefold cross validation on the training data using all features. max = $f(N)$.
3: M =pop the set of features with best average accuracy off S, however randomly choose a feature subset candidate from S every five node expansions.
4: if $f(M) >$ max then
5: max = $f(M)$.
6: else if max has not been changed in $p$ expansions then
7: Go to 15.
```

```
8: end if
9: for all $a_i \in M$ do
10: $M_i = M - a_i$.
11: Run a fivefold cross validation using the $M_i$ subset of features and record the average accuracy $f(M_i)$.
12: Add $M_i$ onto S, which is sorted in ascending order by $f(M_i)$.
13: end for
14: Go to 3.
15: Remove all the elements from S except for the five most accurate feature subsets.
16: if $\mathrm{S} = \emptyset$ then
17: Stop.
18: end if
19: for all elements in S do
20: $\mathrm{M} = \mathrm{pop\ the\ next\ element\ off\ S}$.
21: for all $a_i \in M$ do
22: $M_i = M - a_i$.
23: Run a fivefold cross validation using the $M_i$ subset of features and record the average accuracy $f(M_i)$.
24: Add $M_i$ onto T, which is a queue.
25: end for
26: end for
27: Pick the $q$ most accurate $M_i$ from T and add them to S.
28: Go to 16.
```

After the selection algorithm, we acquired every $B_t(t = 1, 2, \ldots, 29)$, the best average accuracy in fivefold cross validation with t features combination. Then we tested the B_t (the best combination of t features) on the held-out data set and selected the feature subset with least number of features and good accuracy.

VI. Experiments

Several experiments have been done to test our system. The Libsvm [4] SVM software was modified and used in our experiments. Libsvm applies sequential minimal optimization [21]-like decomposition algorithm in its optimization and a one-versus-one approach to do multiclass classification. We modified libsvm to produce a probabilistic output. For comparison, we also implemented a one-versus-all approach. In all experiments the gaussian radial basis function $(k(x, y) = \exp(-\sigma ||x - y||^2))$ was used as the kernel. The parameters C and σ were chosen by fivefold cross validation on each data set.

To evaluate the accuracy of SVMs, we compared with a cascade correlation neural network [10], a C4.5 decision tree with the default pruning settings [19], and two ensembles of decision trees: bagging unpruned decision trees [2] and random forests [3]. There were 100 trees built for each ensemble of decision trees.

A. Initial Experiments

The first training set has a total of 1285 SIPPER images (50 $\mu\mathrm{m}$ resolution), which were selected by marine scientists. It contains images of 64 diatoms, 100 protoctista, 321 doliolids, 366 larvaceans, and 434 trichodesmium. We used C-SVM with parameters $C = 200$ and $\sigma = 0.03$ for one-versus-one and $C = 64$ and $\sigma = 0.08$ for one-versus-all. Table II shows the average accuracy of different learning algorithms from a tenfold cross validation. A paired-t test was used to compare the results at the 95% confidence interval. The SVM one-versus-one approach is significantly more accurate than the other learning algorithms at the 95% confidence level. Also, the running time for one-versus-all and one-versus-one are 2 s and 9 s, respectively on a Pentium 4 PC at 2.6 GHZ. Therefore, the SVM one-versus-one approach outperforms the one-versus-all approach both in accuracy and running time on this data set.

TABLE II
Tenfold Cross Validation Accuracy on the Inital 1285 Image Set

Classifiers	10-fold cross validation accuracy
C4.5 Decision tree	82.2%
Neural network	86.1%
Bagging	87.4%
Random forests	88.2%
SVM (one-vs-all)	86.5%
SVM (one-vs-one)	90.0%

TABLE III
Confusion Matrix of SVM (one-vs-one) From a Tenfold Cross Validation on 1285 SIPPER Images With all 29 Features. P, Di, Do, L, and T Represent Protoctista, Diatom, Doliolid, Larvacean and Trichodesmium, Respectively

	as P	as Di	as Do	as L	as T
P	84.4%	1.6%	9.4%	4.7%	0.0%
Di	2.0%	79.0%	11.0%	6.0%	2.0%
Do	0.8%	0.3%	92.8%	3.1%	0.0%
L	0.8%	0.3%	4.4%	88.0%	6.6%
T	0.0%	0.5%	0.2%	6.2%	93.1%

Table III shows the confusion matrix of the SVM one-versus-one approach from a tenfold cross validation experiment. The overall average accuracy is 90.0%. While we have greater than 84% accuracy on most plankton, we only achieve 79% accuracy on the diatom class. The reason is that we only have 64 diatom samples in our training set and the SVM favors classes with more samples. For instance, assume there is an overlap in the feature space between two classes: one with many examples and one with few examples. It is likely that most examples within that overlap come from the class with more examples. To minimize (8), the decision boundary is pushed away from the class with more examples and thus will favor that class.

B. Experiments With Unidentifiable Particles

The second image set was collected from a deployment of SIPPER in the Gulf of Mexico. A set of 6000 images was selected from the five most abundant types of plankton, which account for 95% of the plankton samples in that run, and manually unrecognizable particles. The five types of plankton are copepods, doliolids, larvaceans, protoctista and trichodesmium. The image quality in this training set is not as good as in the initial experiment. Apart from the shape of image objects, some prior knowledge was used by marine scientists to label the images. Also, we have to classify unidentifiable particles in this experiment.

TABLE IV
TENFOLD CROSS VALIDATION ACCURACY ON THE 6000 IMAGE SET

Classifiers	10-fold cross validation accuracy
C4.5 Decision tree	64.1%
Neural network	70.4%
Bagging	74.2%
Random forest	74.5%
SVM (one-vs-all)	68.7%
SVM (one-vs-one)	75.1%

TABLE V
CONFUSION MATRIX OF SVM (ONE-VS-ONE) FROM A TENFOLD CROSS VALIDATION ON 6000 SIPPER IMAGES WITH ALL 29 FEATURES. C, D, L, P, T, AND U REPRESENT COPOPED, DOLIOLID, LARVACEAN, PROTOCTISTA, TRICHODESMIUM AND UNIDENTIFIABLE PARTICLES, RESPECTIVELY

	As	As	As	As	As	As
	C	D	L	P	T	U
C	84.2%	0.6%	3.1%	1.0%	5.5%	5.6%
D	0.2%	82.9%	2.4%	8.7%	0.4%	5.4%
L	3.2%	1.9%	68.8%	1.4%	11.1%	13.6%
P	1.7%	5.3%	1.1%	84.4%	3.1%	4.4%
T	3.3%	0.6%	9.4%	1.8%	72.5%	12.4%
U	4.3%	3.1%	15.8%	5.4%	13.5%	57.9%

There are a total of 6000 images: 1000 images of each plankton class and 1000 unidentifiable particles. We used C-SVM with $C = 200$ and $\sigma = 0.032$ for one-vs-one and $C = 216$ and $\sigma = 0.114$ for one-vs-all. Table IV shows the average accuracy of different classifiers from tenfold cross validation. A paired-t test was used to compare the results at the 95% confidence interval. The SVM one-versus-one approach is significantly more accurate than all other learning algorithms except the two ensembles of decision trees at the 95% confidence level. Also, the running time for one-vs-all and one-versus-one are 160 s and 610 s, respectively on a Pentium 4 PC at 2.6 GHZ. Therefore, the SVM one-versus-one approach outperforms the one-vsersus-all approach both in accuracy and running time on this data set.

Table V shows the confusion matrix of the SVM one-versus-one approach from a tenfold cross validation. The overall average accuracy is 75.12%. The average accuracy from the five types of plankton is 78.56%.

There are a significant number of larvaceans confused with trichodesmium. This observation disagrees with the first experiment where we had high classification accuracy for both types of plankton. The reason is that some larvacean and trichodesmium are linear objects. Domain experts have prior knowledge of the abundance of larvacean and trichodesmium in some ocean areas. They labeled the linear objects as larvacean or trichodesmium when they know the other plankton were less commonly found in the particular ocean areas examined. Therefore, there are many linear particles without significant features to differentiate between the two types of plankton in this training set, which result in lower classification accuracy on larvaceans and trichodesmium.

It is clear that the one-versus-one approach is superior to the one-versus-all on the two data sets. Therefore, we choose to use one-versus-one approach in our system. We use SVMs to represent SVMs created with the one-versus-one approach by default in the rest of this section.

C. Feature Selection

Feature selection was tested on the larger training set as described in Section VI-B. Although the single SVM seems superior to the other two single classifiers, there is no guarantee that it is still true after feature reduction. Therefore, we experimented with feature selection (wrapper approach) on the SVM and its direct competitor: the cascade correlation neural net. We did not use the decision tree in the comparison because it is far less accurate than the SVM on this data set, thus unlikely to be the best. The data set was randomly divided into two parts: 80% as training and 20% as validation. In this way, we have 1200 data as validation which makes the test result relatively stable and 80% data in training, which is likely to provide a similar feature subset to using all the data. We set the stopping criterion p to be 150 and the beam width q to five in our experiment.

Figs. 8 and 9 show the experimental results of the average accuracy from the fivefold cross validation on the training data and the test accuracy on the validation data respectively. The SVM provided better accuracy than the neural net on both the training set and the validation set when the number of feature is greater than four. To choose the least number of features for the SVM, McNemar's test [9] was applied on the validation set to compute the 95% confidence interval. When the number of features was less than 15, the accuracy would be outside the confidence interval. Therefore, we chose the 15-feature subset as the optimal feature subset and it provided slightly better accuracy than using all the features on the validation data set.

Table VI briefly describes the selected feature subset. A detail description of selected features is contained in attachment 1. Only 1 moment invariant for contour images was selected. It is reasonable because the contours of the plankton images were not stable and was not very helpful in classification. Among the domain specific features, the convex ratio and transparency ratio for images after momphological opening were eliminated. They seem to be redundant to the same features computed on the original images. Therefore, our feature selection approach seems to eliminate irrelevant and redundant features on this image set.

To test the overall effect of feature selection, we applied tenfold cross validation on the whole 6000 image set. The confusion matrix is shown as Table VII. The overall average accuracy is 75.57%. The average accuracy from the five types of plankton is 79.04%. Both indicate that the best 15-feature subset performs slightly better than all 29 features. It is certainly faster to compute the 15 features.

D. Probability Assignment Experiments

In this experiment, we compared our approach (line search for A) and Platt's approach extended to multiple class (gradient descent search for A and B). We used the same training set as in the last experiment with the 15-feature subset. To reduce the overfitting effect from parameter fitting, a threefold cross validation was used to search for the best parameters for both approaches. Since the gradient descent search for A and B is easily stuck in local minima, we varied the initialization several times to get the minimal loss. Table VIII describes the optimal parameters for both approaches. The gradient descent search provided parameters with smaller loss. The line search for a single pa-

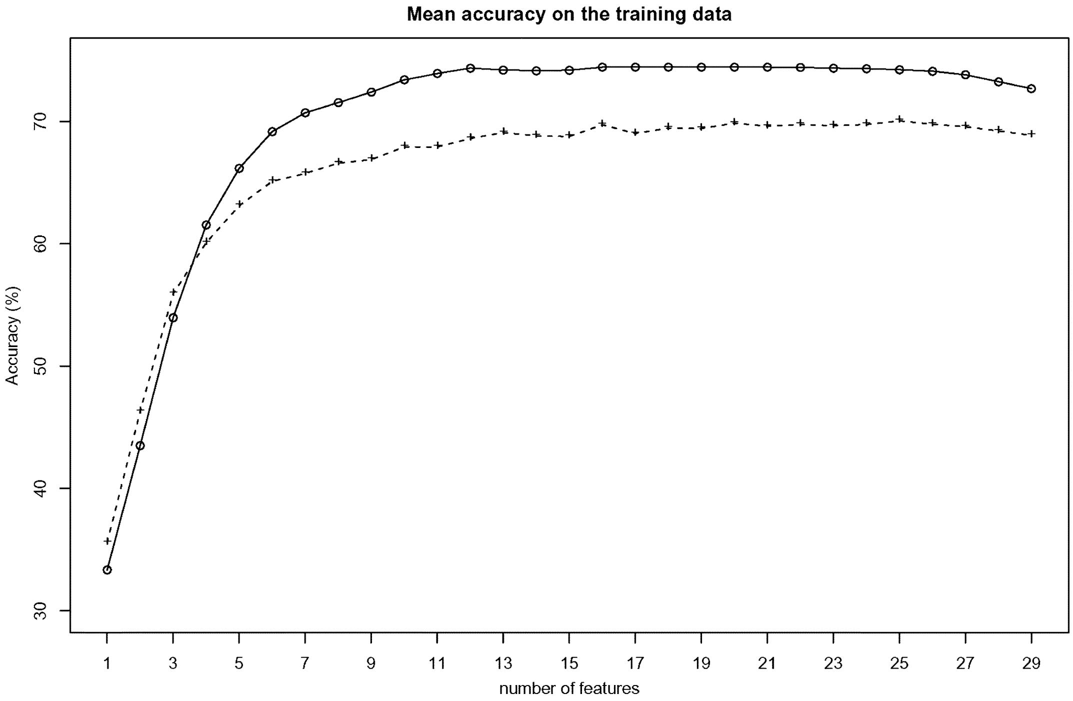


Fig. 8. Feature selection on the training set: The solid line represents accuracy of the SVM and the dashed line represents the accuracy of the neural net.

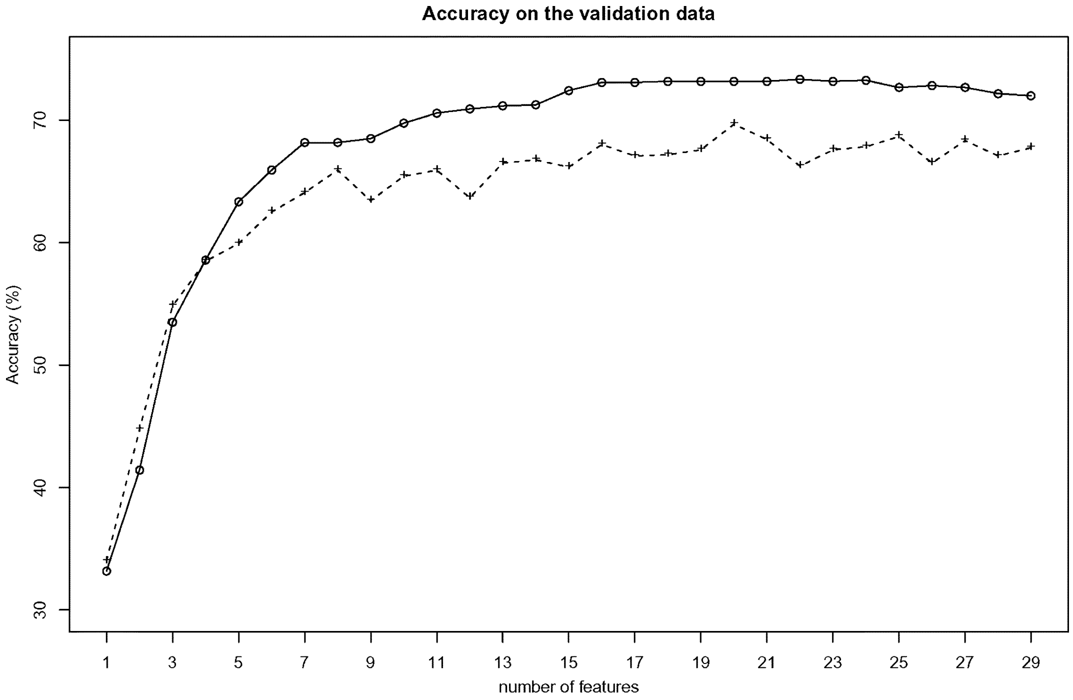


Fig. 9. Selected feature subsets on the validation set: The solid line represents accuracy of the SVM and the dashed line represents the accuracy of the neural net.

TABLE VI
DESCRIPTION OF 15 SELECTED FEATURE SUBSET

Features	Number of original features	Number of selected features
Moment invariants of the original image	7	4
Moment invariants of the contour image	7	1
Granulometric features	7	4
Domain specific features	8	6

rameter A is definitely faster than gradient descent search for A and B with different initializations.

To compare the different parameter sets, we drew a rejection curve from tenfold cross validation using the best parameters for both approaches. The points on the rejection curve were sampled by varying the rejection threshold p, whose range is between 0 and 1. Fig. 10 shows our approach is at least as good as the MLE of A and B. It indicates that $B = 0$ is a reasonable assumption, at least for our data set.

TABLE VII
CONFUSION MATRIX OF SVM (ONE-VS-ONE) FROM A TENFOLD CROSS VALIDATION ON 6000 SIPPER IMAGES WITH THE BEST 15-FEATURE SUBSET. C, D, L, P, T, AND U REPRESENT COPOPED, DOLIOLID, LARVACEAN, PROTOCTISTA, TRICHODESMIUM AND UNIDENTIFIABLE PARTICLES, RESPECTIVELY

	As C	As D	As L	As P	As T	As U
C	84.5%	0.9%	3.1%	0.5%	5.6%	5.4%
D	0.7%	85.2%	1.1%	9.3%	0.4%	3.3%
L	4.3%	2.1%	67.2%	1.1%	12.5%	12.8%
P	1.8%	5.0%	0.7%	85.8%	3.0%	3.7%
T	4.5%	0.4%	10.0%	1.5%	72.5%	11.0%
U	5.1%	2.3%	15.6%	5.4%	13.4%	58.2%

TABLE VIII
BEST PARAMETERS FOR LOG-LIKELIHOOD LOSS FUNCTION

	A	B	-log(L)
Line search for A	87.0	–	4516.5
Gradient descent search for A and B	71.0	0.412	4496.4

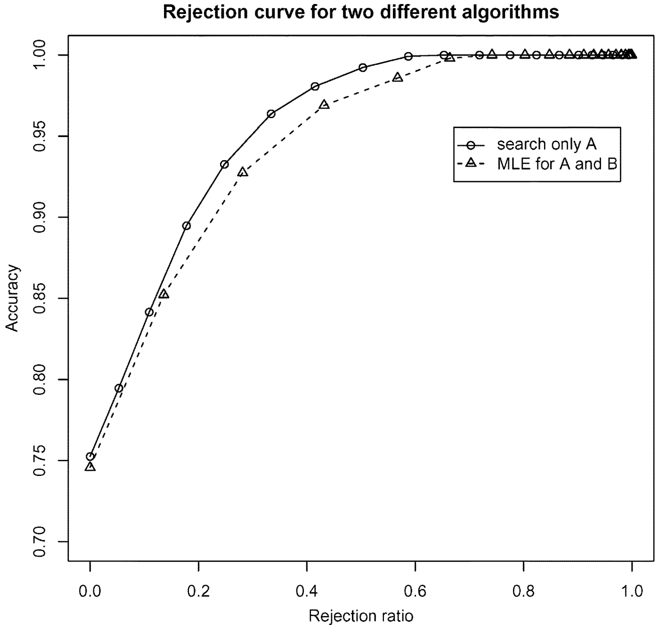


Fig. 10. Rejection curve for both approaches-Overall accuracy versus rejection rate.

VII. CONCLUSIONS AND FUTURE WORK

This paper presents a plankton recognition system for binary SIPPER images. General features, as well as domain specific features, were extracted and a SVM was used to classify examples. We also developed a way to assign a probability value after the multiclass SVM classification. We tested our system on two different data sets. The recognition rate exceeded 90% in one experiment and was over 75% on the more challenging data set with unidentifiable particles. A SVM was more accurate than a C4.5 decision tree and a cascade correlation neural network at the 95% confidence level on the two data sets. The single SVM was significantly more accurate than bagging applied to decision trees and random forests on the smaller data set and was insignificantly more accurate on the larger data set. The wrapper approach with backward elimination successfully reduced the number of features from 29 to 15 and allowed a classifier to be built with slightly better accuracy than using all the features. Our probability model for multiple class SVMs provided a reasonable rejection curve.

The system did not do well at recognizing unidentifiable particles. It is hard to develop specific features to describe the features of the unidentifiable particles because they vary so much. More powerful descriptive and robust general features seem needed in our future work. Recently, an advanced SIPPER system has been developed to produce grayscale SIPPER images at 25 μm resolution. We are in the process of developing methods and features for higher resolution (25 μm resolution) grayscale SIPPER images.

ACKNOWLEDGMENT

The authors would like to acknowledge that the ratio between two eigenvalue and the ratio between head and tail features were suggested by X. Tang (personal communication).

REFERENCES

[1] M. V. Kreveld, M. Overmars, O. Schwarzkopf, and M. d. Berg, *Computational Geometry: Algorithms and Applications*, M. D. Berg, Ed. New York: Springer, 2001.

[2] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[3] ——, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[4] C. C. Chang and C. J. Lin. (2002) LIBSVM: A Library for Support Vector Machines. [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/papers/libsvm.pdf

[5] A. Ciobanu and H. D. Buf, "Identification by contour profiling and legendre polynomials," in *Automatic Diatom Identification*. Singapore: World Scientific, 2002, pp. 167–186.

[6] N. Cristianini and J. Shawe-Taylor, *Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[7] L. F. Costa and R. M. Cesar Jr., *Shape Analysis and Classification*. Boca Raton, FL: CRC, 2001.

[8] T. G. Dietterich, "Machine learning research: Four current directions," *AI Mag.*, vol. 18, no. 4, pp. 97–136, 1997.

[9] ——, "Approximate statistical test for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1924, 1998.

[10] S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," *Adv. Neural Inform. Process. Syst.*, vol. 2, pp. 524–532, 1991.

[11] S. Fischer and H. Bunke, "Identification using classical and new features in combination with decision tree ensembles," in *Automatic Diatom Identification*. Singapore: World Scientific, 2002, pp. 109–140.

[12] D. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*. Cambridge, MA: MIT Press, 2001.

[13] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Networks*, vol. 13, pp. 415–425, Mar. 2002.

[14] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inform. Theory*, vol. IT, no. 8, pp. 179–187, 1962.

[15] J. Kohavi, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1–2, pp. 273–324, 1997.

[16] R. E. Loke and H. d. Buf, "Identification by curvature of convex and concave segment," in *Automatic Diatom Identification*. Singapore: World Scientific, 2002, pp. 141–166.

[17] T. Luo, K. Kramer, D. Goldgof, L. Hall, S. Samson, A. Remsen, and T. Hopkins, "Learning to recognize plankton," in *Proc. IEEE Int. Conf. Systems, Man, Cybernetics*, Oct. 2003, pp. 888–893.

[18] G. Matheron, *Random Sets and Integral Geometry*. New York: Wiley, 1975.

[19] J. R. Quinlan, *C4.5: Programs from Empirical Learning*. San Mateo, CA: Morgan Kaufmann, 1993.

[20] I. Pitas, *Digital Image Processing Algorithms and Applications*. New York: Wiley, 2000.

[21] J. Platt, "Fast training of support vector machines using sequential minimal optimization," *Advances in Kernel Methods – Support Vector Learning*, pp. 185–208, 1999.

[22] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classifiers*, pp. 61–74, 2000.

[23] S. Samson, T. Hopkins, A. Remsen, L. Langebrake, T. Sutton, and J. Patten, "A system for high resolution zooplankton imaging," *IEEE J. Ocean. Eng.*, vol. 26, pp. 671–676, Oct. 2001.

[24] L. M. Santos and H. D. Buf, "Identification by Gabor features," in *Automatic Diatom Identification*. Singapore: World Scientific, 2002, pp. 187–220.

[25] X. Tang, W. K. Stewart, L. Vincent, H. Huang, M. Marra, S. M. Gallager, and C. S. Davis, "Automatic plankton image recognition," *Artif. Intell. Rev.*, vol. 12, no. 1–3, pp. 177–199, 1998.

[26] X. Tang, F. Lin, S. Samson, and A. Remsen, "Binary plankton image classification," , submitted for publication.

[27] V. N. Vapnik, *The Nature of Statistical Learning Theory*: Springer, 2000.

[28] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, "Feature selection for SVMs," *Neural Inform. Proc. Syst.*, vol. 13, 2001.

[29] M. H. F. Wilkinson, A. C. Jalba, E. R. Urbach, and J. B. T. M. Roerdink, "Identification by mathematical morphology," in *Automatic Diatom Identification*. Singapore: World Scientific, 2002, pp. 221–244.

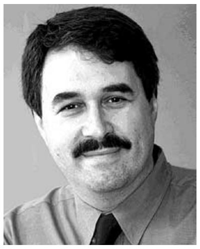
[30] C. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curve," *IEEE Trans. Comput.*, vol. 21, pp. 269–281, 1972.

Tong Luo received the B.S. degree in electrical engineering from Tsinghua Universtiy, Beijing, China, and the M.Eng. degree in control theory and system from the Institute of Automation, Beijing, Chinese Academy of Sciences, China in 1997 and 2000, respectively. He is currently pursuing the Ph.D degree in computer science and engineering at the University of South Florida, Tampa.

His research interests are in machine learning, data mining, and pattern recognition.

Kurt Kramer (M'99) received the B.S. degree in computer engineering from the University of South Florida, Tampa, in 2002 and he is currently pursuing the M.S. degree in computer science at the same university.

His research interests are in the area in of image processing/recognition and AI. He is currently working on automatic recognition of plankton.



Dmitry B. Goldgof (S'84–M'87–SM'93) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 1989.

He is currently a Professor in the Department of Computer Science and Engineering at the University of South Florida, Tampa, and a Member of the H. Lee Moffitt Cancer Center. His research interests include motion analysis, computer vision, image processing and its biomedical applications, pattern recognition, and bioinformatics. He has graduated ten Ph.D. degree students and 27 M.S. degree students, and has published over 50 journal publications, over 100 conference publications, 16 books chapters, and four books. He is actively involved with a number of professional conferences.

Dr. Goldgof is the North American Editor for Image and Vision Computing Journal and Associate Editor for IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS, B.

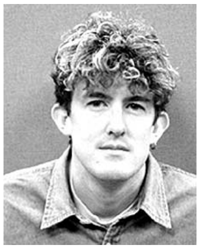
Lawrence O. Hall (S'85–M'85–SM'98–F'02) received the B.S. degree in applied mathematics from the Florida Institute of Technology, Melbourne, and the Ph.D. degree in computer science from Florida State University, Tallahassee, in 1980 and 1986, respectively.

He is currently a Professor of computer science and engineering at University of South Florida, Tampa. His research interests lie in distributed machine learning, data mining, pattern recognition, and integrating AI into image processing. The exploitation of imprecision with the use of fuzzy logic in pattern recognition, the AI and learning is a research theme. He has authored over 150 publications in journals, conferences, and books. Recent publications appear in *Artificial Intelligence in Medicine*, *Pattern Recognition Letters*, JAIR, IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, the International Conference on Data Mining, the Multiple Classifier Systems Workshop, and the European Conference on Machine Learning. He co-edited the 2001 joint North American Fuzzy Information Processing Society (NAFIPS) IFSA conference proceedings.

Dr. Hall received the IEEE SMC Society Outstanding contribution award in 2000. He is Past President of NAFIPS, Vice President for Membership of the SMC society, the Editor-In-Chief of the IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS, B, and also, Associate Editor for IEEE TRANSACTIONS ON FUZZY SYSTEMS, *International Journal of Intelligent Data Analysis*, and *International Journal of Approximate Reasoning*.

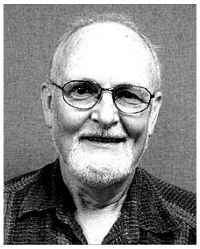
Scott Samson received the B.S.E degree in biomedical engineering, and the Ph.D. degree in electrical engineering, both from the University of Iowa, Iowa City, in 1990 and 1994, respectively.

He is currently an Optoelectronics Engineer at the Center for Ocean Technology, University of South Florida, Tampa. He has been involved in the design of the SIPPER underwater particle imaging instrument, visible and ultraviolet optical sensors, and microelectromechanical (MEMS) components, and has co-authored several papers and presentations on these topics. In addition to his imaging and MEMS experience, his expertise includes integrated optical component development, as well as analog and digital electronics design.



Andrew Remsen received the B.S. degree in zoology from the University of Wisconsin, Madison, in 1991, and is currently pursuing the Ph.D. degree in biological oceanography at the College of Marine Science, University of South Florida, Tampa.

He has been involved in the development of SIPPER from its inception and is currently responsible for field operations of the instrument. His research interests include, oceanographic sensor development, zooplankton ecology, marine food web interactions, and remote sensing.



Thomas Hopkins is a Biological Oceanographer and Adjunct Professor in the College of Marine Science, the University of South Florida, Tampa. Most of his research effort is centered on the structure of the pelagic food web in Antarctic and Gulf of Mexico ecosystems with emphasis on zooplankton and micronekton predator-prey relationships. Current research involves leading a program to develop optical sensors for counting, sizing, imaging, and determining the spatial distribution of plankton.