

Automatic plankton image recognition with co-occurrence matrices and Support Vector Machine

Qiao Hu^{1,*}, Cabell Davis²

¹Department of Applied Ocean Physics and Engineering, and ²Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543-1541, USA

ABSTRACT: A long-standing problem in plankton ecology is sparseness of taxa-specific data. New optical imaging systems are becoming available which can acquire high-resolution data on the abundance and biomass of plankton taxa. The Video Plankton Recorder (VPR) has been designed and used for automatic sampling and visualization of major planktonic taxa at sea in real time, providing high-resolution data over a broad range of scales. Although these optical systems produce digital images of plankton that can be automatically identified by computer, the limited accuracy of automatic classification methods can reduce confidence in subsequent abundance estimates, especially in areas where a taxon is in low relative abundance. This paper describes an improved classification system for automatic identification of plankton taxa from digital images. Classifiers are trained from a set of images that were classified by human experts. The data set used to verify the classification system contains over 20 000 planktonic images manually sorted into 7 different categories. The new method uses co-occurrence matrices (COM) as the feature, and a Support Vector Machine (SVM) as the classifier. This new method is compared against a previous plankton recognition system, which used moment invariants, Fourier descriptors and granulometry as features and a learning vector quantization neural network as a classifier. The new method reduced the classification error rate from 39 to 28 %. Subsequent plankton abundance estimates are improved by more than 50 % in regions of low relative abundance. In general, the reduction in classification error was due to a combination of the use of COM and SVM.

KEY WORDS: Pattern recognition · Video Plankton Recorder · Co-occurrence matrix · Support Vector Machine

—Resale or republication not permitted without written consent of the publisher—

INTRODUCTION

Plankton distributions are patchy and require high-resolution sampling for adequate quantification (Fasham 1978, Haury et al. 1978, Omori & Ikeda 1984, Mackas et al. 1985). Historically, a large effort has been devoted to developing various zooplankton sampling devices (Wiebe & Benfield 2003). Over the last decade, new optical imaging devices have been developed that are allowing high-resolution distributions of plankton taxa to be determined (Video Plankton Recorder, VPR, Davis et al. 1992a,b; Underwater Video Profiler, Gorsky et al. 1992, 2000; Ichthyoplankton Recorder,

Froese et al. 1990, Lenz et al. 1995; FlowCam, Sieracki et al. 1998; *in situ* video, Tiselius 1998; Shadow Image Particle Profiling Evaluation Recorder, SIPPER, Samson et al. 2001). The VPR has been used to identify planktonic taxa in real time (Davis et al. 2004, 2005), measure distributions nearly continuously over a broad range of scales (Davis et al. 1996, Gallagher et al. 1996), and quantify abundance and biomass of delicate plankton and particulate matter (Davis et al. 1992b, Norrbin et al. 1996, Ashjian et al. 2001, in press a, in press b). Initial data from the VPR and other optical samplers required manual processing, a tedious and time-consuming effort. For example, VPR surveys typ-

*Email: qhu@whoi.edu

ically generate 10^4 to 10^6 images per day (10^5 to 10^7 per cruise) that need to be sorted to taxa, making manual analysis impractical.

Research on automatic plankton classification has been on-going for many years (Jeffries et al. 1984, Rolke & Lenz 1984, Berman et al. 1990, Hofstra et al. 1994, Costas et al. 1995). Early systems worked on images from well-controlled laboratory conditions, and had not been applied to field-collected images. More recently, artificial neural networks have come to play a central role in classifying plankton images (Simpson et al. 1991, Boddy & Morris 1993, Culverhouse et al. 1994, 2003, Tang 1998, Tang et al. 1998, Toth & Culverhouse 1999). The datasets used to develop and test these classifiers were fairly small (Tang et al. 1998, Culverhouse et al. 2003). Furthermore, distinctive images were chosen to train and test the classifier, while the classifier needs to classify all the images from the field, including difficult ones and those that cannot be identified even by a human expert. The accuracy reported for a classifier built only from distinctive images will in general be much higher than its 'true' performance when the classifier is applied to all field data (Davis et al. 2004).

The features used in the early systems were mostly shape-based. Jeffries et al. (1984) used moment invariants as well as Fourier descriptors and morphometric relations as features. These features worked quite well under well-defined laboratory imaging conditions. The recognition rate reported by Jeffries was 90% on 6 taxonomic groups, but the system required significant human interaction and was not suitable for *in situ* applications. Initial automatic identification of VPR images was performed using the method described in Tang et al. (1998) which introduced granulometry curves (Vincent 1993), along with traditional features such as moment invariants, and Fourier descriptors and morphometric measurements, as features in automatic plankton image recognition. This method used a Learning Vector Quantization (LVQ) neural network for the classifier (Tang 1998) and achieved 92% classification accuracy on a subset of VPR images for 6 taxonomic groups. In developing this method, the image subset included distinctive images for both training and testing the classifier. Classification accuracy for the entire set of images was much lower. Davis et al. (2004) analyzed this method in detail using a large set of field-collected images and determined that the average classification performance on the whole dataset was 61%. The performance disagreement between previous methods (Jeffries et al. 1984, Tang et al. 1998) and that described in Davis et al. (2004) is due to the nature of field-captured images. Unlike well-controlled laboratory conditions, field images are

often occluded (objects truncated at edge of image), and shape-based features such as moment invariants and Fourier descriptors are very sensitive to occlusion. In addition, a significant number of field-collected images are unidentifiable by a human expert due to object orientation and position in the image volume¹. These unidentifiable images were not used in our initial training and testing subsets (Tang et al. 1998, although occluded images were included). A recent study by Luo et al. (2004) showed that including unidentifiable objects lowered the recognition rate from 90 to 75% for their dataset from the SIPPER. Other work has shown that it is important to include 'other' (Davis et al. 2004) or 'reject' (Grosjean et al. 2004) categories to better estimate abundance.

Nonlinear illumination of images also makes perfect segmentation (binarization) impossible, even with background brightness gradient correction. Due to grayscale gradient, the same object can have a different segmented shape depending where the object is in the field-of-view, thus causing shape-based features to be less reliable. Another type of feature we can calculate from the grayscale images is texture-based. Due to the early success of shape-based features on plankton images from well-controlled laboratory imaging conditions, texture-based features have not been widely used in plankton image recognition. In this paper, we describe a method for classification of plankton images using co-occurrence matrices as features and a Support Vector Machine as the classifier. We call this method COM-SVM. As co-occurrence matrices extract localized features from the images, based on pixel brightness, it is less affected by occlusion, illumination gradients and projection variance. These characteristics make COM-SVM more suitable for classification of field-collected images of plankton.

MATERIALS AND METHODS

VPR system overview. The VPR is a towed underwater video microscope that images plankton and seston in the size range of 100 μm to 1 cm (a full description is given in Davis et al. 1992a, 1996, 2004, 2005). The VPR also includes a suite of physical sensors including a CTD, fluorometer, optical backscatter sensor, PAR sensor and flowmeter. Video images are sent to a focus detection system that performs brightness

¹Objects can be hard to identify due to their position in the image volume. If part of the object is out of this volume, the resulting image will be occluded. Nonlinear illumination makes objects from the dark region more likely to be occluded by global segmentation, a problem correctable by background gradient removal (Davis et al. 2005)

correction², segmentation, labeling, size thresholding, edge detection, coalescing and region-of-interest (ROI) generation. Each ROI is saved as a Tagged Image Format (TIF) file using time of day in milliseconds as the file name (Davis et al. 1996, 2004). A subset of these files is identified (labeled) by a human expert and is used to train a classifier. After training, all the files generated by the focus detection system are automatically classified by the classifier. The classification results are used to estimate the abundance of the taxon of interest.

Co-occurrence matrices. Spatial gray level co-occurrence provides second-order statistics from the images. Julesz (1962) first used first-order and second-order statistics in texture discrimination. The co-occurrence method was first proposed by Haralick et al. (1973) as a texture feature and it has been widely used thereafter. It is based on the estimation of the joint probability distribution of pairs of pixels with gray level i and j , spatial distance d and angle θ in an image. Each element in the co-occurrence matrix is the occurrence of pairs of pixels having gray levels i and j , and a certain spatial relationship in the whole image (i.e. distance d and angle θ). Thus, for an image of L quantization level, the size of its co-occurrence matrix is $L \times L$. The number of co-occurrence matrices is dependent on the number of different separation distances and quantized levels of angle. For computation efficiency, the angle is usually quantized to 45 or 90°. It is hard to select d without prior information. It is common to choose $d = 1$. In our study, we quantized the angle to 45°, which ended up with 4 different angles (0, 45, 90, 135°), and chose $d = 1, 4, 8, 16$ pixels.

Support Vector Machine. The Support Vector Machine (SVM) was proposed by Vapnik (1995, 1998) and has yielded excellent results in a variety of data classification tasks. It is primarily a 2-class classifier and involves 2 steps. First, the feature vectors $\mathbf{x}$ of the training samples are mapped into a high (potentially infinite) dimensional space $\mathcal{H}$. A hyperplane is then constructed in order to separate the training samples in $\mathcal{H}$. Different mappings $\mathbf{x} \mapsto \Phi(\mathbf{x}) \in \mathcal{H}$ construct different SVMs.

The mapping $\Phi(\cdot)$ is performed by a kernel function $K(\cdot, \cdot)$ which defines an inner (dot) product in $\mathcal{H}$. The decision function (i.e. the hyperplane), f , given by an SVM is:

$$f(\mathbf{x}) = \text{sign}[\langle \mathbf{w}, \Phi(\mathbf{x}) \rangle + b] = \text{sign}\left[\sum_{i=1}^m \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right] \quad (1)$$

where $\mathbf{w}$ and b define the orientation and translation of f , respectively, i is the training sample index, y is the class label, and α is a scalar.

The goal in training a SVM is to find the separating hyperplane which has the maximal distance to the closest training samples in space $\mathcal{H}$. This distance is called the margin. These particular training feature vectors that are used to determine optimal hyperplanes are called support vectors. In order to cope with non-separable cases, a set of slack variables $\xi_i \geq 0$ are introduced. If there are m training samples: $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ with class label $y_i \in \{-1, 1\}$, the classification reduces down to the following optimization problem:

$$\text{minimize } L_P(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{m} \sum_{i=1}^m \xi_i \quad (2)$$

with relaxed separation constraints:

$$y_i [\langle \Phi(\mathbf{x}_i), \mathbf{w} \rangle + b] \geq 1 - \xi_i, \quad i = 1, \dots, m \quad (3)$$

where $\mathbf{w}$ is normal to the hyperplane and C is a scalar value that controls the trade-off between the empirical risk and margin width. The dual formulation is usually easy to solve, and is defined as:

$$\text{maximize } L_D(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (4)$$

subject to constraints:

$$\sum_{i=1}^m \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, m \quad (5)$$

There are 3 main ways to extend SVMs from 2-class to multi-class classification: (1) The simplest is the one-versus-all approach (Rifkin & Klautan 2004) in which a set of binary SVMs are trained to separate one class from the rest. The main drawback of this approach is that the sample size is unbalanced, with the number of images in the selected class typically much less than that containing the rest. (2) Another method is Error-Correcting Output Codes (Dietterich & Bakiri 1995) in which a series of binary problems are generated from a multi-class problem by splitting the original set of classes into 2 subsets. This method appears promising but is untested for plankton image data. (3) In the present study, we used a pair-wise approach (cf. Luo et al. 2004), where all possible pairs of 2 classes were used to build binary SVMs. For the classification with n classes, $n(n-1)/2$ binary SVMs are needed. This yields 21 binary SVMs for our case of 7 classes (see Table 1).

An important property of the SVM is that the complexity of the classifier is characterized by the number of support vectors instead of the dimension of the hyperspace $\mathcal{H}$. As a result, the SVM is less prone to over-fitting than other methods (Vapnik 1995, Burges 1998, Duda et al. 2001).

Confusion matrix. A confusion matrix contains information about actual and predicted counts done by a classification system. Performance of classification systems can be easily evaluated from the matrix. For example, in Table 1, out of 5983 copepods (column sum), 4401 of them were classified as copepods. Likewise, 97

²This correction does not affect ROI grayscale values. Original ROI grayscale values are used for feature calculation

Table 1. Confusion matrix for EN302, VPR Tow 7, based on the co-occurrence matrix classifier using the hold-out method. Column and row headings are coded as: C1, copepod; C2, rod-shaped diatom chains; C3, *Chaetoceros* chains; C4, *Chaetoceros socialis* colonies; C5, hydroid medusae; C6, marine snow; C7, 'other'. True counts (i.e. human counts) for a given taxon are given in the columns, while counts by automatic identification (i.e. computer counts) are given in the rows. Correct identifications by the computer are given along the main diagonal, while the off-diagonal entries are incorrect identification by the computer. The unit of the entries is count except for the last which is in percent. Overall, the accuracy for this classifier was 72 %

	C1	C2	C3	C4	C5	C6	C7
C1	4401	73	8	6	79	22	1078
C2	97	3326	11	10	17	54	329
C3	63	30	391	44	148	47	236
C4	58	61	13	2098	46	267	390
C5	232	10	29	15	1293	50	168
C6	68	50	15	148	151	1142	136
C7	1064	106	11	73	62	68	1967
Accuracy	74	91	82	88	72	69	46

copepods were classified as rod-shaped diatom chains. The row labeled as 'Accuracy' is the probability of detection for each category, which measures the probability that the classification system will label correctly label for each category given the object belongs to that category. Another important number calculated from the confusion matrix is false alarm. It is defined as the probability that the classifier will label the object as one group when the object does not belong to that group.

Working data set. Our working data set was obtained from a 24 h VPR tow (VPR-7) in the Great South Channel off Cape Cod, Massachusetts, during June 1997 on the RV 'Endeavor'. The VPR was towed from the ship in an undulating mode, forming a tow-yo pattern between the surface to near bottom. The images were taken by the high magnification camera, which has an image volume of 0.5 ml. The total sampled volume³ during the deployment was about 2.6 m³. There were over 20 000 images captured during this tow. All the images were manually identified (labeled) by a human expert into 7 major categories (including one category as 'other', comprising rare taxa and unidentifiable objects). These are the most abundant categories in this area. In this tow, about 21 % of the images belonged to the 'other' category. Most of these images were unidentifiable by human experts. The rest were rare species including coil-shaped diatom chains, ctenophores, chaetognaths, poly-

chaetes and copepod nauplii (see Davis et al. 2004). The manual identification took weeks, while the COM-SVM classified the same amount of data in 2 to 3 h on a 1 GHz personal computer. The representative samples (images) are shown in Fig. 1. We treated manual labels (identifications) as the ground truth for comparing different classifiers.

Feature extraction and classification. Each of the image files was first quantized to 16 grayscale levels. The co-occurrence matrices were calculated from 4 different angles (0, 45, 90, 135°) and 4 different distances (1, 4, 8, 16 pixels). A frequency normalization was performed by dividing each entry in the co-occurrence matrices by the total number of neighbor pairs. For example, for an image of size $M \times N$, when the relationship between nearest horizontal neighbors is ($d = 1, \theta = 0^\circ$), there will be a total of $2N(M - 1)$ nearest horizontal neighbor pairs. For every 4 matrices with the same distance, the mean and range matrices were calculated. Thus, for each image, 8 co-occurrence matrices were computed. The energy, contrast, correlation, variance, inverse-difference moment, and entropy of these matrices (Haralick et al. 1973) were calculated and used as feature vector elements. These features were further normalized to have zero mean and unit standard deviation.

The Ohio State University (OSU) support vector machine (OSU-SVM) was used to classify these feature vectors. The OSU-SVM was developed by J. Ma, Y. Zhao, and S. Anhalt for the Matlab platform using Chang and Lin's LIBSVM algorithm⁴. It is available at http://www/ece.osu.edu/~maj/osu_svm. The OSU-SVM uses decomposition in its optimization and a pairwise approach to do multi-class classification. Different kernels were tested on our data set. In our experiment, the Gaussian radial basis function (RBF) performed best in terms of validation error. The Gaussian RBF kernel is defined as:

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (6)$$

where σ is a scalar value.

Two data subsets were randomly picked from our working data set. These data sets had 200 samples per taxon and were used to train and validate the SVM classifier, respectively. The whole data set was used to make the confusion matrices. Values of σ and the regularization constant C were optimized based on the classification error found from validation data set. Values of $\sigma = 0.1$ and $C = 50$ gave the best classifier performance. Since the validation data set was used to tune the classifier parameters, it is not valid to use them to test the classifier (generate confusion matrix).

³As pointed out in Davis et al. (2005), although the volume imaged by VPR is small compared to the volume filtered by a plankton net, the VPR can still provide an equivalent or better estimate of plankton abundance

⁴Software available at www.csie.ntu.edu.tw/~cjlin/libsvm

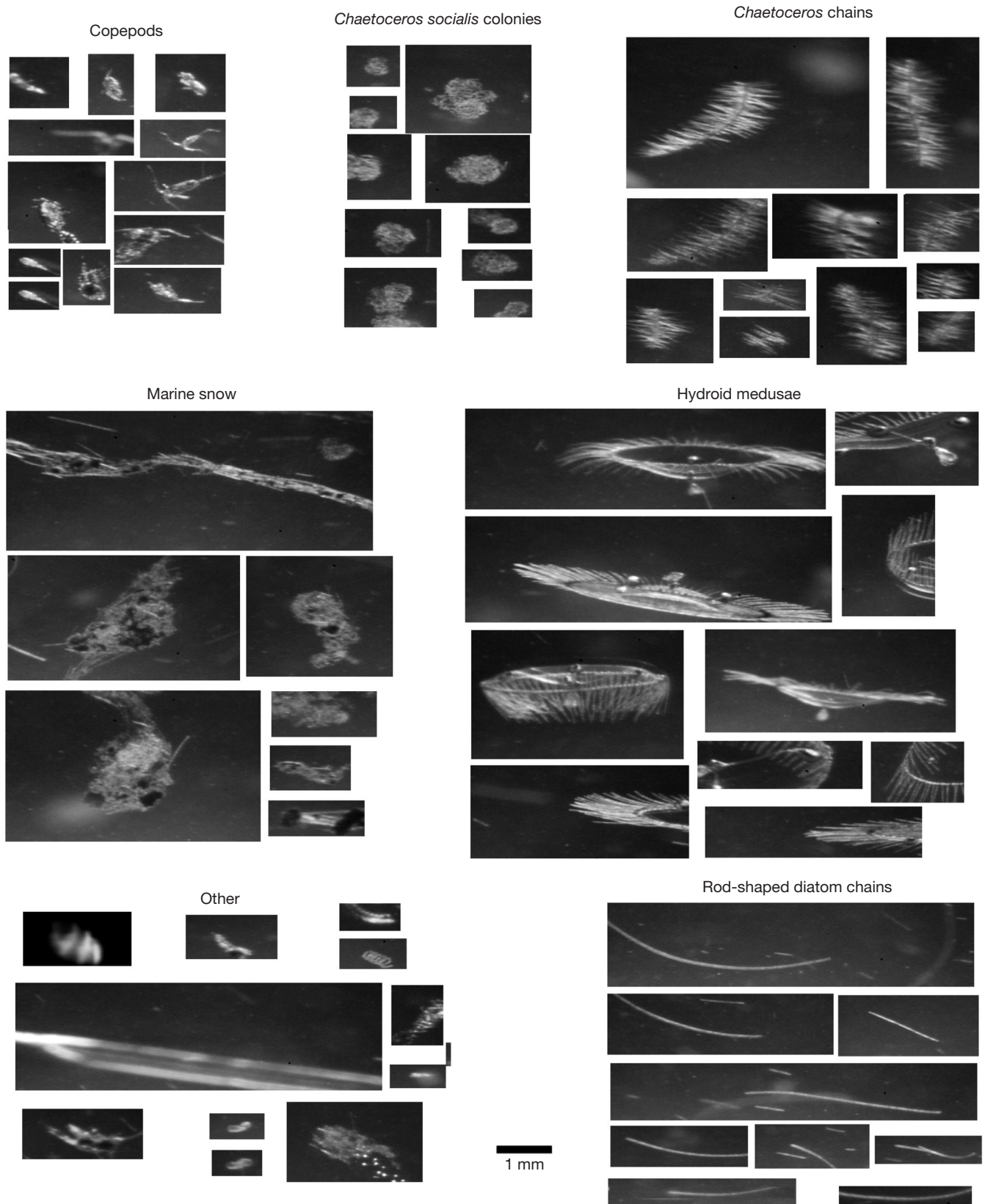


Fig. 1. Example VPR images from the study site on Georges Bank (RV 'Endeavor' cruise EN302). Representative images of each taxon are shown. Image width ranges from 1 to 7 mm

RESULTS AND DISCUSSION

We compared the performance of our COM-SVM to our prior plankton classifier (Tang et al. 1998, Davis et al. 2004). The COM-SVM yielded a 28% reduction in recognition error rate (cf. Tables 1 & 2, respectively)⁵. The overall performance of the COM-SVM was 72% compared to 61% for the previous system. The COM-SVM classifier performed better than the original features (combined moment invariants, Fourier descriptors and granulometry curve) and the neural network (OF-NN) classifier for almost all the categories except the 'other' (cf. Tables 1 & 2). This finding supports our idea that for field-collected samples, texture-based features are more important than global shape-based features in plankton classification, due to occlusion, nonlinear illumination, or projection variance inherent in images captured *in situ* by existing imaging systems. Most occlusions occur when part of an organism is out of the image volume. Some occlusions happen when part of an organism is darker than the rest because of the nonlinear illumination. The global segmentation only segments part of the organism. The situation of nonlinear illumination should be improved using a proposed ring-illuminator in future instruments. A small amount of occlusion can also happen when an out-of-focus organism is in the light path of an in-focus organism. This situation only occurs when the concentration of the plankton is very high (≥ 10 ind. ml⁻¹).

Although Culverhouse et al. (2003) showed that human experts were far from perfect for certain difficult classification tasks such as plankton identification, for simplicity, we considered the human expert as a 'perfect classifier' in this study. The effect of training with contaminated samples is a very interesting research topic. Research on handwritten characters by Schölkopf & Smola (2002) suggests that classifier accuracy was not too sensitive to a small amount of contamination. Further study is needed to decide how 'clean' the training set needs to be to have a reliable classifier.

Testing the effects of different kernels and their parameters revealed that the SVM classifier was robust to both kernel function type and parameters specific to the kernel (cf. Table 3). For radial basis function (Gaussian kernel), the recognition rate was not sensitive to the choice of penalty constant C . For a wide range of C (10 to 500), the recognition rate only changed by 2%. The recognition rate was more sensitive to the kernel width σ for the radial basis function; however, the recognition rate was still fairly constant over a wide range of σ . For the polynomial kernel, the

Table 2. Mean confusion matrix for EN302, VPR Tow 7, based on the learning vector quantization method neural network classifiers built with different randomly selected sets of 200 training ROIs using the hold-out method (from Davis et al. 2004). Column and row headings are as in Table 1. True counts (i.e. human counts) for a given taxon are given in the columns, while counts by automatic identification (i.e. computer counts) are given in the rows. The correct identifications by the computer are given along the main diagonal, while the off-diagonal entries are incorrect identification by the computer. The unit of the entries is count except for the last row which is in percent. Overall, the accuracy for this classifier was 61%

	C1	C2	C3	C4	C5	C6	C7
C1	3604	482	26	29	104	95	1048
C2	155	2822	2	18	59	25	231
C3	138	20	325	181	281	184	212
C4	127	26	26	1757	143	181	302
C5	72	40	13	29	829	186	97
C6	116	37	34	119	214	758	165
C7	1771	224	47	247	151	135	2185
Accuracy	60	77	68	73	46	48	51

recognition rate changed from 69 to 74% with a change in polynomial order from 1 to 6. For the sigmoid kernel, the change in classifier performance was relatively small, and the performance itself was similar to that obtained using the other kernels. Among all the kernel methods, the top performances differed by only 1%. The similarity between these different classifiers in performance improvement indicates that classification is not sensitive to the classifiers being used. Specifically, the sigmoid kernel SVM is equivalent to certain types of NN classifier, implying that COM features are more relevant than shape-based features to the plankton classification problem.

Table 3. Performance of the classifier with different kernel widths (σ), regulation penalties (C) and kernel types, where d is polynomial degree and κ is kernel coefficient. The recognition rate on the independent test set is shown

Gaussian kernel ($C = 50$) $k(x, x') = \exp\left(\frac{-\ x - x'\ ^2}{2\sigma^2}\right)$							
σ	0.2	0.5	1	2	3	5	10
Rate	63	69	72	75	75	73	70
Gaussian kernel ($\sigma = 2$) $k(x, x') = \exp\left(\frac{-\ x - x'\ ^2}{2\sigma^2}\right)$							
C	10	20	50	100	200	500	
Rate	73	75	75	75	74	73	
Polynomial kernel $C = 5$ $k(x, x') = \langle x, x' \rangle^d$							
d	1	2	3	4	5	6	
Rate	74	74	74	72	69	69	
Sigmoid kernel $k(x, x') = \tanh(\kappa \langle x, x' \rangle + \theta)$							
κ	0.01	0.02	0.05	0.1	0.2	0.5	
Rate	69	71	73	74	72	60	

⁵The difference of true counts of 2 tables is due to the original NN classifier ignoring organisms whose detected outline is less than a certain number of pixels

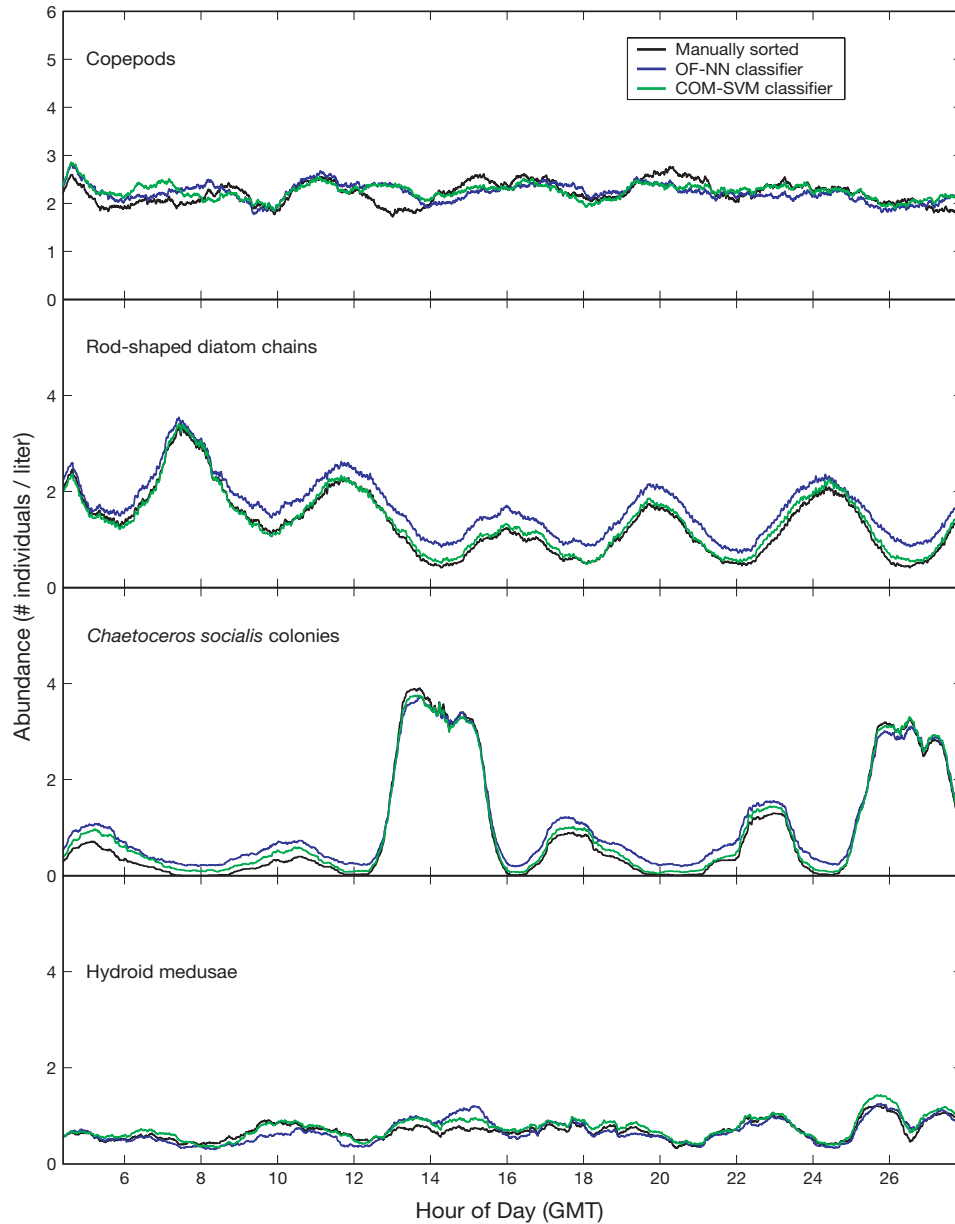


Fig. 2. Comparison of 2 automated classifiers with human expert classified results for 4 dominant classes along the tow timescale. The data are first binned in 10 s time intervals. A 1 h smoothing window is applied to the binned data

In estimating plankton abundance, the performance of COM-SVM was uniformly better than the OF-NN classifier (Fig. 2). Abundance estimates for both classifiers had the same trends as the manually sorted result. Differences in abundance between these methods, quantified using the Kullback-Leibler (KL) distance method (Duda et al. 2001) for all 4 taxa, revealed a closer agreement between COM-SVM and manually sorted than between manually sorted and OF-NN (Table 4), reflect-

Table 4. Kullback-Leibler (KL) distance (D_{KL}) estimation for difference in abundance between COM-SVM and manually sorted, and between OF-NN and manually sorted. The KL distance is dimensionless. For 2 identical abundance curves, the KL distance is 0, while for 2 random distributions, the KL distance is 0.5. Note lower values of COM-SVM than OF-NN for all 4 taxa

	Copepod	Rod-shaped diatom	<i>Chaetoceros socialis</i>	Hydroid medusae
$D_{KL}(\text{COM-SVM, hand})$	0.0036	0.0022	0.0225	0.0048
$D_{KL}(\text{OF-NN, hand})$	0.0041	0.0113	0.0742	0.0188

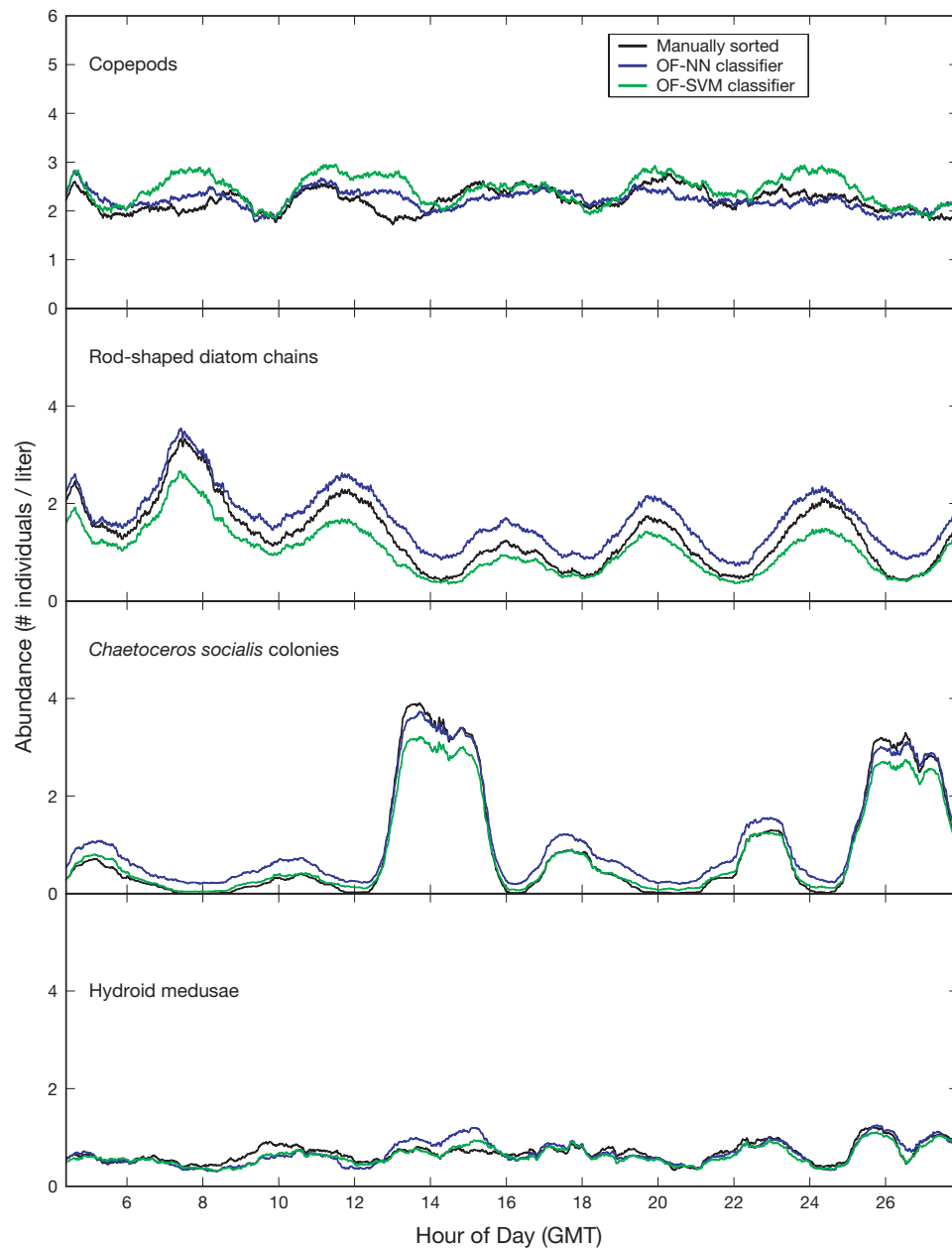


Fig. 3. Same plot as Fig. 2 except COM-SVM was replaced by OF-SVM

ing the higher accuracy of the COM-SVM method. In order to investigate the relative contribution of COM and SVM in improving performance, the SVM classifier was trained using original features. The abundance estimation of this classifier (OF-SVM) was compared to that of the original classifier (OF-NN). The OF-SVM classifier was found to perform better than the OF-NN classifier in regions of low abundance for *Chaetoceros socialis* colonies. On the other hand, the OF-SVM classifier gave underestimates in relatively high abundance regions. In overall performance, the OF-SVM classifier and OF-NN classifier were fairly

similar (Fig. 3). As discussed elsewhere (Davis et al. 2004), when the relative abundance of a taxon is above 20 to 25%, the abundance estimation error due to misclassification falls well within the natural variation for replicate plankton tows. In areas of low relative abundance, the accuracy of the abundance estimates is typically much lower due to a relatively high false alarm rate (Solow et al. 2001, Davis et al. 2004). The 28% reduction in recognition error results in a reduction in abundance estimate error rate by more than 50% for *C. socialis* colonies in areas of low relative abundance (Fig. 4). The reduction in abundance error rate is due

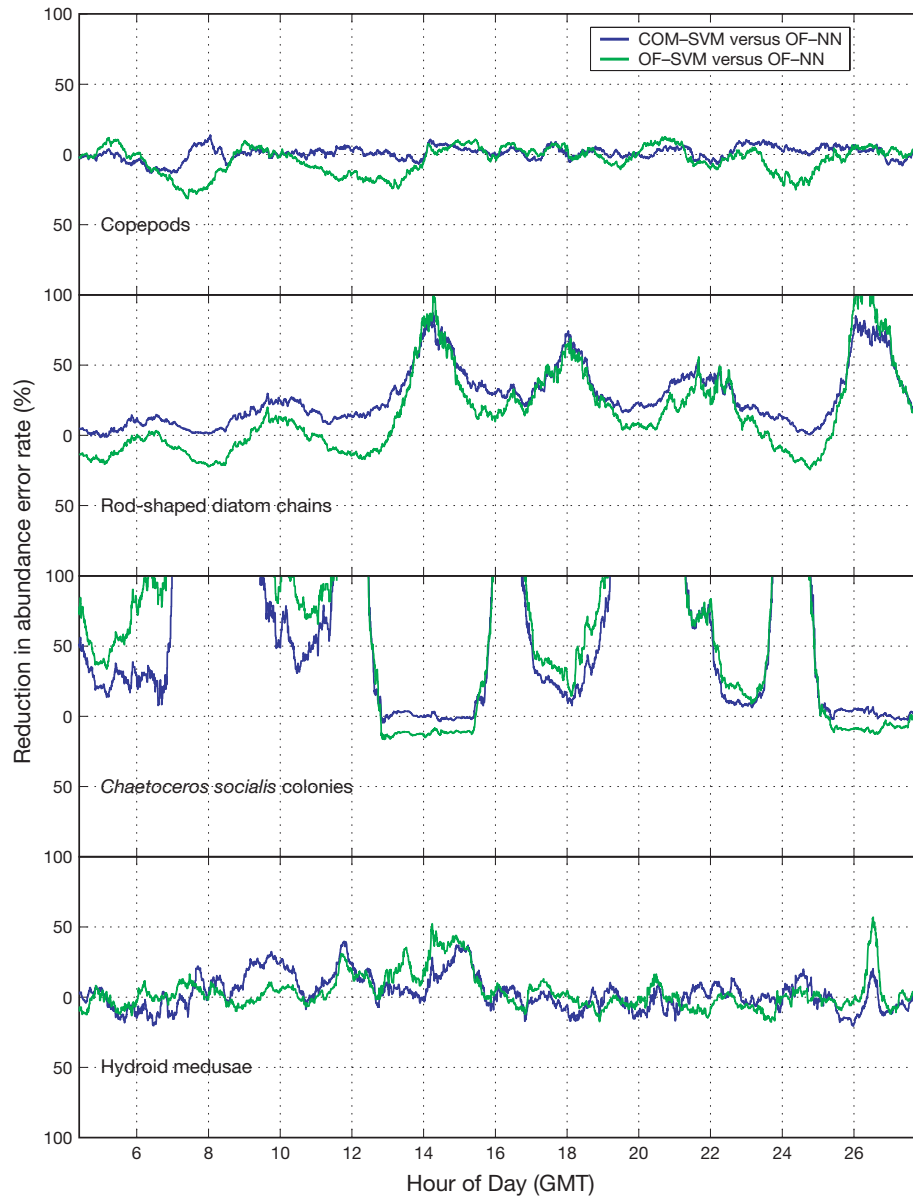


Fig. 4. Reduction in the relative abundance estimation error rate between COM-SVM and OF-NN, and between OF-SVM and OF-NN. The positive value indicates that COM-SVM/OF-SVM is better than OF-NN. The negative value indicates COM-SVM/OF-NN is worse than OF-NN

to the use of both COM and SVM (Fig. 4). Positive values indicate improved performance, while negative values indicate worse performance. COM-SVM outperformed OF-SVM in most cases, except in regions of low *C. socialis* abundance. The latter performance difference was due to a general underestimation by the OF-SVM classifier and a consequent increase in *C. socialis* abundance in these regions. These observations further support the idea that the use of texture-based features (i.e. co-occurrence matrix) is the main reason for performance improvement in our classification system.

The pair-wise approach was chosen in order to extend the binary SVM classifier to the multi-class SVM classifier used in this study. Another approach using the Error-Correcting Output Coding method (Dietterich & Bakiri 1995) also appears very promising and is becoming an active research topic (Allwein et al. 2001, Crammer & Singer 2002, Passerini et al. 2004). Further analysis of this method is the subject of future study.

The new COM-SVM method only uses texture-based features (i.e. co-occurrence matrices) to automatically classify plankton images. Shape-based features also carry a substantial amount of information that can be

used for classification. An attempt was made to directly stack texture-based features and shape-based features into a single feature vector, and train the classifier on this single feature vector (with and without principal component analysis). Only a very limited improvement (less than 1 %) in recognition rate was obtained. This method of combining features was only one approach, and further research is needed to determine whether other methods (such as weighting each individual feature by its discriminative power) for combining features can yield improved identification accuracy.

Given the growing trend towards optical imaging of marine biota, new methods of automatic identification are needed to improve classification accuracy. The texture-based method presented here can be used for a wide-variety of image classification problems, since it is not sensitive to occlusion and lighting gradients and is independent of shape-based features.

CONCLUSION

In this paper, we used a texture-based feature, co-occurrence matrices, to classify plankton images taken in the field using the VPR. This method yielded a 72 % overall recognition rate compared to 61 % for a previous recognition system that used shaped-based features. Shape-based features are the primary ones currently used in automatic plankton recognition systems due to their early success with plankton images taken in the laboratory. Texture-based features were found to work better for field-collected images of plankton because they are less sensitive to occlusion, non-uniform lighting and projection variance.

SVM was used to train the classifier. Classifier performance was not sensitive to kernel type or to the exact parameter values used for specific kernels. From prior work, we know that selection of representative training samples is an important factor. In order to accurately assess classifier performance, a random set of training samples from the field is recommended.

Multi-scale texture features are captured with multiple separation distances. Scale invariance is achieved by normalization of co-occurrence matrices. Rotation invariance is achieved by using only the range and mean co-occurrence matrices.

Continued improvements in accuracy of automatic image recognition methods will enable wider use of this powerful approach. The growing use of underwater optical imaging methods requires more emphasis on development and improvement of new automatic identification techniques.

The method described here is a step toward the long term goal of highly-accurate automatic identification of plankton from optical imaging systems.

Acknowledgements. The work was funded by the Office of Naval Research grant N00014-98-10032 and National Science Foundation grants OCE-9820099 and OCE-0000580. We appreciate comments by 3 anonymous reviewers.

LITERATURE CITED

- Allwein EL, Schapire RE, Singer Y (2001) Reducing multiclass to binary: a unifying approach for margin classifiers. *J Mach Learn Res* 1:113–141
- Ashjian CJ, Davis CS, Gallagher SM, Alatalo P (2001) Distribution of plankton, particles and hydrographic features across Georges Bank described using the Video Plankton Recorder. *Deep-Sea Res II* 48:245–282
- Ashjian CJ, Gallagher SM, Plourde S (in press a) Transport of plankton and particles between the Chukchi and Beaufort Seas during summer 2002, described using a Video Plankton Recorder. *Deep-Sea Res II*
- Ashjian CJ, Davis CS, Gallagher SM, Alatalo P (in press b) Characterization of the zooplankton community and size composition in relation to hydrography and circulation in the Sea of Japan. *Deep-Sea Res II*
- Berman MS, Katsinis C, Jeffries HP, Lambert RM (1990) Image analysis techniques for the identification and measurement of marine zooplankton. *EOS Trans Am Geophys Union* 71:94 (Abstract)
- Boddy L, Morris CW (1993) Analysis of flow cytometry data—a neural network approach. *Comput Microb* 5:17–22
- Burges CJC (1998) A tutorial on support vector machines for pattern recognition. *Data Min Knowl Disc* 2:121–167
- Costas E, San Andres MI, Gonzalas-Gil S, Aguilera A, Lopez-Rodas V (1995) A procedure to estimate okadaic acid in whole dinoflagellate cells using immunological techniques. *J Appl Phycol* 7:407–411
- Crammer K, Singer Y (2002) On the learnability and design of output codes for multiclass problems. *Mach Learn* 47: 201–233
- Culverhouse PF, Ellis RE, Simpson R, Williams R, Pierce RW, Turner JT (1994) Categorisation of five species of *Cymatocylis* (Tintinidae) by artificial neural network. *Mar Ecol Prog Ser* 107:273–280
- Culverhouse PF, Williams R, Reguera B, Herry V, Gonzalez-Gill S (2003) Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Mar Ecol Prog Ser* 247:17–25
- Davis CS, Gallagher SM, Berman MS, Haurly LR, Strickler JR (1992a) The Video Plankton Recorder (VPR): design and initial results. *Arch Hydrobiol Beih Ergebn Limnol* 36:67–81
- Davis CS, Gallagher SM, Solow AR (1992b) Microaggregations of oceanic plankton observed by towed video microscopy. *Science* 257:230–232
- Davis CS, Gallagher SM, Marra M, Stewart WK (1996) Rapid visualization of plankton abundance and taxonomic composition using the Video Plankton Recorder. *Deep-Sea Res II* 43:1947–1970
- Davis CS, Hu Q, Gallagher SM, Tang X, Ashjian CJ (2004) Real-time observation of taxa-specific plankton distributions: an optical sampling method. *Mar Ecol Prog Ser* 284:77–96
- Davis CS, Thwaites F, Gallagher SM, Hu Q (2005) A three-axis fast-tow digital Video Plankton Recorder for rapid surveys of plankton taxa and hydrography. *Limnol Oceanogr Meth* 3:59–74
- Dietterich TG, Bakiri G (1995) Solving multiclass learning problems via error-correcting output codes. *J Artif Intell Res* 2:263–286

- Duda RO, Hart PE, Stork DG (2001) Pattern classification. John Wiley & Sons, New York
- Fasham MJR (1978) The statistical and mathematical analysis of plankton patchiness. *Oceanogr Mar Biol Annu Rev* 16:43–79
- Froese R, Barthel KG, Welsch W, Rolke M and 5 others (1990) Development of an underwater video system for recording of ichthyoplankton and zooplankton. *ICES CM* 90:6
- Gallager SM, Davis CS, Epstein AW, Solow A, Beardsley RC (1996) High-resolution observations of plankton spatial distributions correlated with hydrography in the Great South Channel, Georges Bank. *Deep-Sea Res II* 43:1627–1664
- Gorsky G, Aldorf C, Picheral M, Kage M, Garcia Y, Favole J (1992) Vertical distribution of suspended aggregates determined by a new Underwater Video Profiler. *Ann Inst Oceanogr* 68(1–2):275–280
- Gorsky G, Picheral M, Stemann L (2000) Use of the Underwater Video Profiler for the study of aggregate dynamics in the north Mediterranean. *Estuar Coast Shelf Sci* 50(1): 121–128
- Grosjean P, Picheral M, Warembourg C, Gorsky G (2004) Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *ICES J Mar Sci* 61:518–525
- Haralick RM, Shanmugan K, Dinstein I (1973) Texture feature for image classification. *IEEE T Syst Man Cyb* 3:610–621
- Haury LR, McGowan JA, Wiebe PH (1978) Patterns and processes in the time-space scales of plankton distributions. In: Steele JH (ed) *Spatial patterns in plankton communities*. Plenum Press, New York, p 277–327
- Hofstraat JW, van Zeijl WJM, de Vreeze MEJ, Peeters JCH, Peperzak L, Colijn F, Rademaker TWM (1994) Phytoplankton monitoring by flow cytometry. *J Plankton Res* 16:1197–1224
- Jeffries HP, Berman MS, Poularikas AD, Katsinis C, Melas I, Sherman K, Bivins L (1984) Automated sizing, counting and identification of zooplankton by pattern recognition. *Mar Biol* 78:329–334
- Julesz B (1962) Visual pattern discrimination. *IRE Trans Inform Theory* 8:84–92
- Lenz J, Schnack D, Petersen D, Kreikemeier J, Hermann B, Mees S, Wieland K (1995) The Ichthyoplankton Recorder: a video recording system for in situ studies of small-scale plankton distribution patterns. *ICES J Mar Sci/J Cons Int Explor Mer* 52(3–4): 409–417
- Luo T, Kramer K, Goldgof DB, Hall LO, Samson S, Remsen A, Hopkins T (2004) Recognizing plankton images from the shadow image particle profiling evaluation recorder. *IEEE T Syst Man Cy B* 34:1753–1762
- Mackas DL, Deman KL, Abbott MR (1985) Plankton patchiness: biology in the physical vernacular. *Bull Mar Sci* 37:652–674
- Norrbin MF, Davis CS, Gallager SM (1996) Differences in fine-scale structure and composition of zooplankton between mixed and stratified regions of Georges Bank. *Deep-Sea Res II* 43:1905–1924
- Omori M, Ikeda T (1984) *Methods in marine zooplankton ecology*. John Wiley & Sons, New York
- Passerini A, Pontil M, Frasconi P (2004) New results on error correcting output codes of kernel machines. *IEEE T Neural Networ* 15:45–54
- Rifkin R, Klautan A (2004) In defense of one-vs-all classification. *J Mach Learn Res* 5:101–141
- Rolke M, Lenz J (1984) Size structure analysis of zooplankton samples by means of an automated image analyzing system. *J Plankton Res* 6:637–645
- Samson S, Hopkins T, Remsen A, Langebrake L, Sutton T, Patten J (2001) A system for high-resolution zooplankton imaging. *IEEE J Oceanic Eng* 26:671–676
- Schölkopf B, Smola AJ (2002) *Learning with kernels*. MIT Press, Cambridge, MA
- Sieracki CK, Sieracki ME, Yentsch CS (1998) An imaging in-flow system for automated analysis of marine microplankton. *Mar Ecol Prog Ser* 168:285–296
- Simpson R, Culverhouse PF, Ellis R, Williams R (1991) Classification of *Euceratium* Gran. in neural networks. *IEEE International Conference on Neural Networks in Ocean Engineering*, Washington, DC, USA, August 1991. IEEE, Piscataway, NJ, p 223–230
- Solow A, Davis CS, Hu Q (2001) Estimating the composition of a sample when individuals are classified with error. *Mar Ecol Prog Ser* 216:309–311
- Tang X (1998) Multiple competitive learning network fusion for object classification. *IEEE T Syst Man Cy B* 28:532–543
- Tang X, Stewart WK, Vincent L, Huang H, Marra M, Gallager SM, Davis CS (1998) Automatic plankton image recognition. *Artif Intell Rev* 12:177–199
- Tiselius P (1998) An *in situ* video camera for plankton studies: design and preliminary observations. *Mar Ecol Prog Ser* 164:293–299
- Toth L, Culverhouse PF (1999) Three-dimensional object categorisation from static 2D views using multiple coarse channels. *Image Vision Comput* 17:845–858
- Vapnik V (1995) *The nature of Statistical Learning Theory*. Springer-Verlag, New York
- Vapnik V (1998) *Statistical Learning Theory*. John Wiley & Sons, New York
- Vincent L (1993) Morphological grayscale reconstruction in image analysis: application and efficient algorithm. *IEEE T Image Process* 2:176–201
- Wiebe PH, Benfield MC (2003) From the Hensen net toward 4-dimensional biological oceanography. *Prog Oceanogr* 56:7–136

Editorial responsibility: Kenneth Sherman (Contributing Editor), Narragansett, Rhode Island, USA

Submitted: September 7, 2004; Accepted: February 14, 2005
Proofs received from author(s): June 1, 2005