

Automatic Plankton Image Recognition

XIAOOU TANG¹, W. KENNETH STEWART¹, LUC VINCENT²,
HE HUANG¹, MARTY MARRA³, SCOTT M. GALLAGER¹ and
CABELL S. DAVIS¹

¹*Woods Hole Oceanographic Institution, Challenger Drive, MS #7, Woods Hole, MA 02543-1108, USA (Voice: 508-289-3226; Fax: 508-457-2191; Email: xtang@whoi.edu);*

²*Xerox Imaging Systems, 9 Centennial Drive, Peabody, MA 01960, USA;*

³*Vexcel Corporation, 2477 55th Street, Boulder, CO 80301, USA*

Abstract. Plankton form the base of the food chain in the ocean and are fundamental to marine ecosystem dynamics. The rapid mapping of plankton abundance together with taxonomic and size composition is very important for ocean environmental research, but difficult or impossible to accomplish using traditional techniques. In this paper, we present a new pattern recognition system to classify large numbers of plankton images detected in real time by the Video Plankton Recorder (VPR), a towed underwater video microscope system. The difficulty of such classification is compounded because: 1) underwater images are typically very noisy, 2) many plankton objects are in partial occlusion, 3) the objects are deformable and 4) images are projection variant, i.e., the images are video records of three-dimensional objects in arbitrary positions and orientations. Our approach combines traditional invariant moment features and Fourier boundary descriptors with gray-scale morphological granulometries to form a feature vector capturing both shape and texture information of plankton images. With an improved learning vector quantization network classifier, we achieve 95% classification accuracy on six plankton taxa taken from nearly 2,000 images. This result is comparable with what a trained biologist can achieve by using conventional manual techniques, making possible for the first time a fully automated, at sea-approach to real-time mapping of plankton populations.

1. Introduction

Plankton form the base of the food chain in the ocean and are a fundamental component of marine ecosystem dynamics. Understanding the ecological and physical processes controlling population dynamics of plankton over a wide range of scales, from centimeters to hundreds of kilometers, is essential for understanding how climate change and human activities affect marine ecosystems. Such studies require large-scale, high-resolution mapping of plankton abundance and taxonomic and size composition. High-resolution temporal sampling is needed to measure tidal, diel, and seasonal variability of population abundance and composition. Until recently, however, it has been difficult or impossible to conduct such extensive sampling because plankton abundance is highly variable in time and space and cannot be quantified with sufficient resolution using conventional sampling methods.

Traditionally, plankton surveys are conducted with such equipment as towed nets, pumps, and Niskin bottles. Because of the laborious deployment process and limited sample storage space on ship, the spatial sampling rate is extremely low. The painstaking and error-prone post-processing – manual counting of samples through a microscope and data entry – may take months or years, which effectively prohibits large-scale, high-resolution, three-dimensional surveys over periods of time. However, accurate estimates of production and growth can be made only if the interactions of organisms with one another and with the local environment are estimated from samples drawn at appropriate intervals of space and time (Owen 1989).

To help overcome the limitations of traditional plankton sampling instruments, a new Video Plankton Recorder (VPR) has been developed (Davis et al. 1992; Davis et al. 1992). As the VPR is towed through the water, it continuously captures magnified plankton images, providing a spatial resolution of plankton distribution on scales from microns to over 100 km. The amount of image data collected over even short periods of time can be overwhelming, necessitating an automated approach to plankton recognition. This approach would not only save a great deal of man power, but also make the real-time sorting of plankton possible. Real-time abundance and distribution data on zooplankton and accompanying environmental variables are needed to guide researchers during field studies on population and community processes, just as physical oceanographers have for decades used real-time measurements of temperature and conductivity to adjust their survey strategy according to observed phenomena (Paffenhofer 1991).

Now that high-quality images of individual plankton can be obtained with the VPR, our approach to the full automation of at-sea analysis of plankton size and taxonomic composition focuses on the development of an image analysis and pattern recognition system for real-time processing of the large volume of image data being acquired. Our development approach includes three parts (Davis et al. 1992; Davis et al. 1996): 1) a hardware/software system for preprocessing of the images (including real-time image capture, object detection, and in-focus analysis) and digital storage of detected object images; 2) pattern recognition algorithms for automated identification and classification of planktonic taxa; 3) incorporation of the pattern recognition algorithms into a high-performance image analysis system to achieve a real-time processing capability. Development of a preprocessing and acquisition system as described in Step 1 has been completed and used to detect and save subimages of planktonic taxa in real-time while at sea (Davis et al. 1996).

In this paper, we mainly address Step 2 and demonstrate an automated approach to plankton image classification. Our experimental data sets differ from those used for most previous pattern recognition researches in four

aspects: 1) the underwater images are much noisier, 2) many objects are in partial occlusion, 3) the objects are deformable, and 4) images are projection variant, i.e., the images are video records of three-dimensional objects in arbitrary positions and orientations. Figure 1, which shows example sub-images extracted from the larger video fields, illustrates the diversity of images within individual taxa.

By combining granulometric features with such traditional two-dimensional shape features as moment invariants and Fourier boundary descriptors, we extract a more complete description of the plankton patterns. Then, using an improved Learning Vector Quantization (LVQ) neural network classifier, we classify the plankton images into several taxonomic categories. The algorithms are tested on six classes taken from nearly 2,000 plankton images. The resulting classification accuracy is comparable with what a trained biologist can achieve using traditional manual techniques.

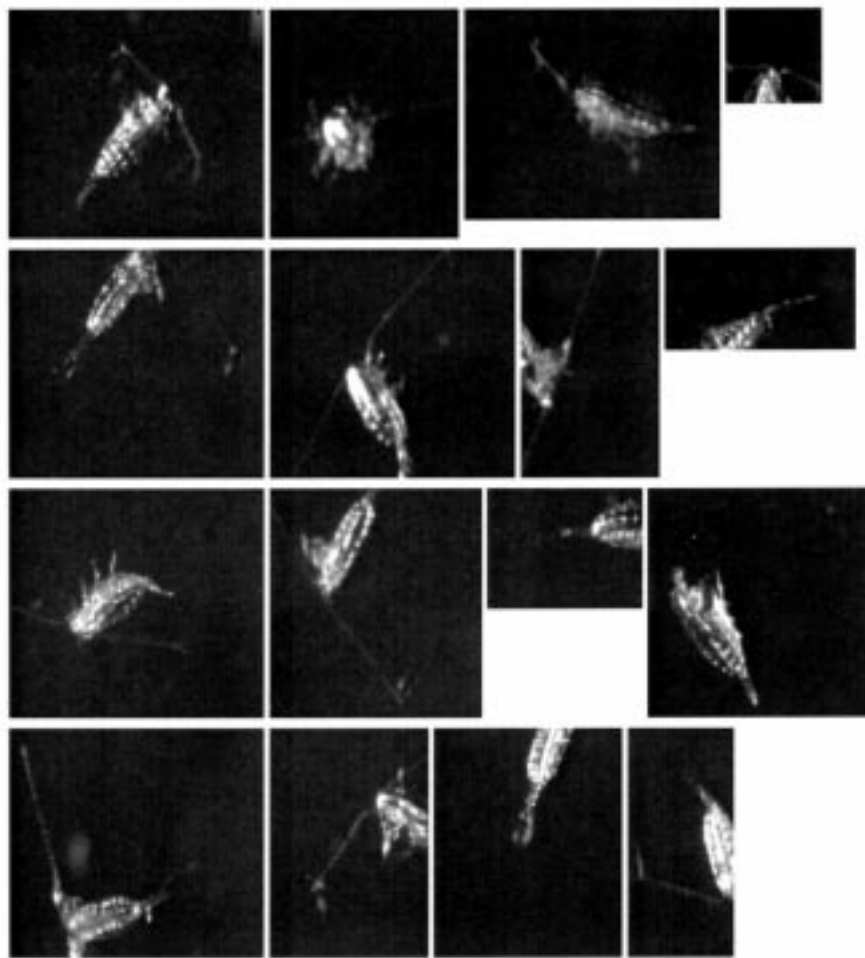
The paper is organized as follows. In Section 2, the three feature extraction methods – moment invariants, Fourier boundary descriptors, and granulometric features – are described, along with a feature selection algorithm. We then introduce an improved LVQ classifier. Section 3 describes real-time data acquisition and image processing. In Section 4, experimental results from the classification of the six plankton taxa are reported. We summarize our conclusions and point to future work in Section 5.

2. Methodology

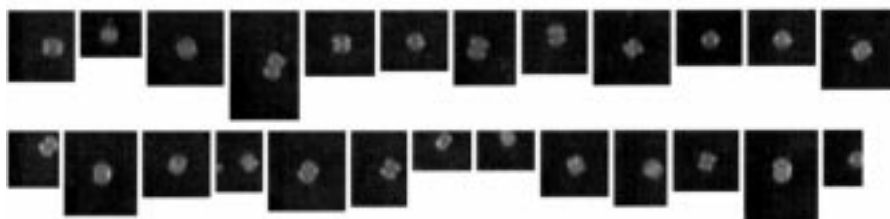
2.1. Feature extraction

Developing algorithms for classification of two-dimensional shapes insensitive to position, size, and orientation is an important problem in pattern recognition. Application of these algorithms range from industrial inspection and scene analysis to optical character recognition. The most widely used shape features are moment invariants and Fourier boundary descriptors. Classification of three-dimensional projection-variant objects is even more difficult.

In this paper, we introduce gray-scale granulometric features as a powerful pattern descriptor, which captures both shape and texture signatures, as a step toward addressing the three-dimensional problem. We also combine the three types of feature vectors to form a more complete description of the plankton patterns. We briefly review the three feature types then describe an effective feature selection method.

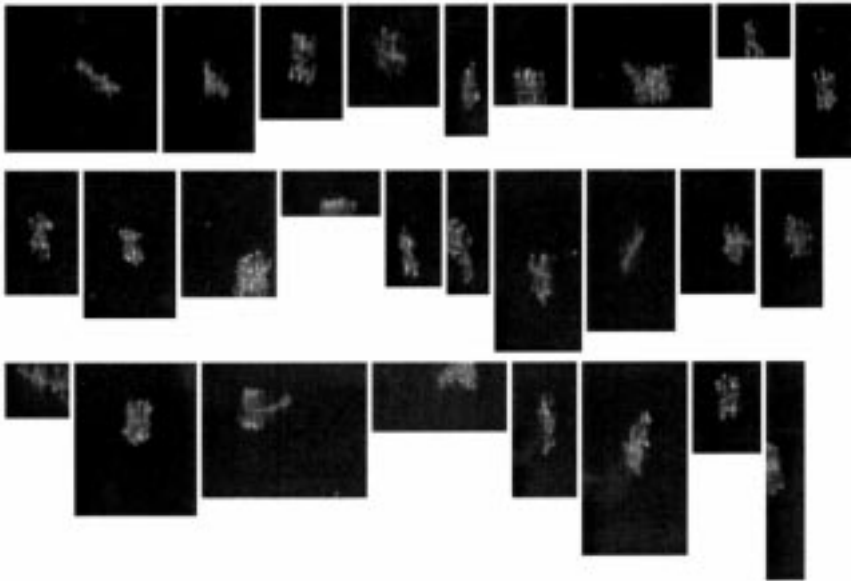


(a) CALANUS



(b) DIAT-CENTR

Figure 1(a-f). Sample images for each of the six types of plankton.

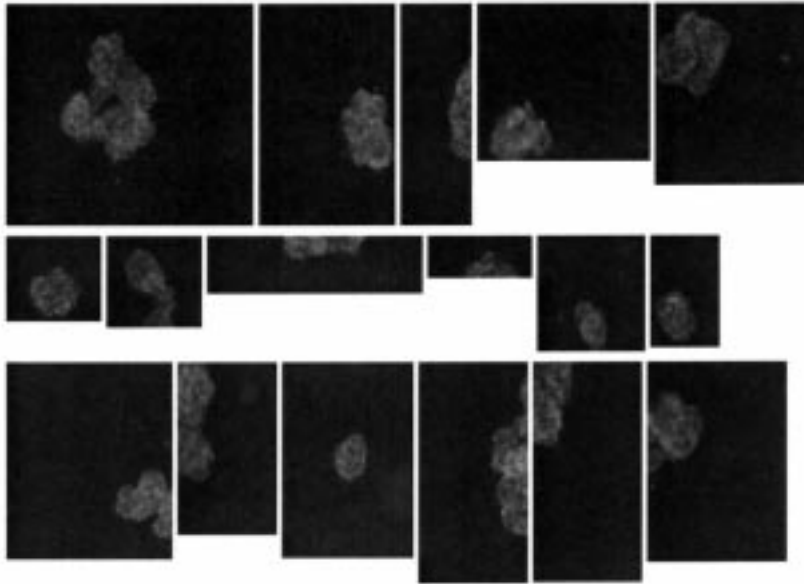


(c) DIAT-CHAET

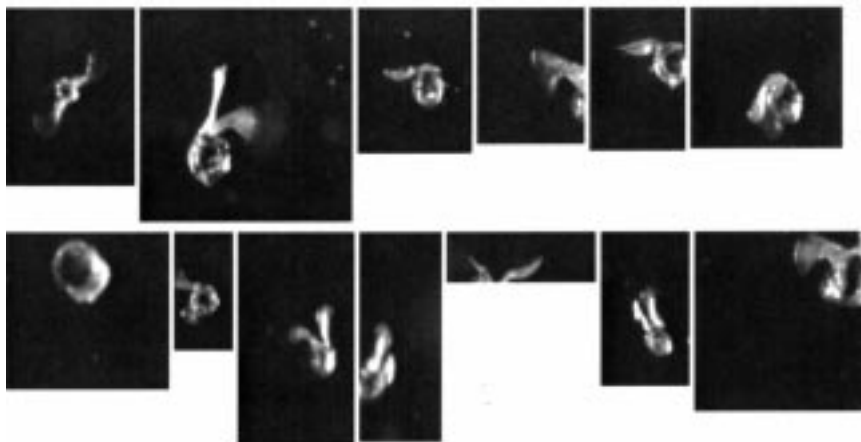


(d) DIATOM

Figure 1(a-f). Continued.



(e) DIATOMCOLO



(f) PTEROPOD

Figure 1(a-f). Continued.

2.1.1. Moment invariants

The concept of moments as invariant image features was first introduced by Hu (1962), and later revised by Reiss (1991). Moments and functions of moments have been used as pattern features in many applications. Some examples and comparisons of different features are found in Gonzalez and Wintz (1987), Reeves et al. (1988) and Teh and Chin (1991). In our experiments, we use the seven invariant moments described by Hu (1962), and Gonzalez (1987).

A $(p + q)$ th order moment of a continuous image function $f(x, y)$ is defined as

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots \quad (1)$$

For digital images the integrals are replaced by summations and m_{pq} becomes

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y). \quad (2)$$

The central moments of a digital image can be expressed as

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y), \quad (3)$$

where

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}}.$$

The normalized central moments are derived from these central moments as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \gamma = \frac{p+q}{2} + 1, \quad p+q = 2, 3, \dots \quad (4)$$

Based on methods of algebraic invariants, Hu (1962) derived seven invariant moments $\phi_i, i = 1 \dots 7$, using nonlinear combinations of the second and third normalized central moments. These invariant moments possess the desirable properties of being translation, rotation, and scale invariant.

2.1.2. Fourier descriptor

The first in depth study of the Fourier descriptor was given by Zahn and Roskies (1972) and later refined by Persoon and Fu (1977). Recently, more research effort has been devoted to shape classification based on Fourier descriptors (Reeves et al. 1988; Kauppinen et al. 1995; Reti and Czinege 1989). The most often used boundary models include the curvature function, centroidal radius, and complex contour coordinates. Kauppinen et al. (1995)

give a detailed experimental comparison of different models. We use the radius Fourier Descriptor (FD) and complex contour Fourier descriptor, which were shown to be the best among FDs tested in Kauppinen et al. (1995).

Consider a closed boundary defined by a closed sequence of successive boundary pixel coordinates (x_i, y_i) . The centroidal radius function expresses the distance of boundary points from the centroid (x_c, y_c) of the object,

$$r_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}. \quad (5)$$

A complex contour coordinate function is simply the coordinates of the boundary pixels in an object centered coordinate system represented as complex numbers,

$$z_i = (x_i - x_c) + j(y_i - y_c). \quad (6)$$

Since both functions are computed around the centroid of the object, they are automatically translation invariant. To achieve rotation and scale invariance, a Fourier transformation of the boundary signature is generally used. For digital images we use the discrete Fourier transform (DFT). The shift invariant DFT magnitude gives a rotation invariant feature vector. Scale invariance is accomplished by normalizing all DFT magnitudes by the DFT magnitude of the zero frequency or the fundamental frequency component.

The feature vector for a radius Fourier Descriptor is

$$FD_r = \left[\frac{|F_1|}{|F_0|} \dots \frac{|F_{N/2}|}{|F_0|} \right], \quad (7)$$

where N is the boundary function length and F_i denotes the i th component of the Fourier spectrum. Notice that only half of the spectrum need to be used because of the symmetric property of the Fourier transform of real functions.

The contour Fourier method transforms the complex coordinate function in Equation (6) directly. The feature vector is

$$FD_c = \left[\frac{|F_{-(N/2-1)}|}{|F_1|} \dots \frac{|F_{-1}|}{|F_1|} \frac{|F_2|}{|F_1|} \dots \frac{|F_{N/2}|}{|F_1|} \right]. \quad (8)$$

In this case, both positive and negative halves of the complex spectrum are retained. Because the complex function is centered around the coordinate origin, F_0 has zero value and F_1 is used for the normalization.

2.1.3. Granulometric features

The above traditional features are mostly developed in a well controlled pattern environment. Test images are usually shifted, scaled, and rotated

versions of a small set of perfect images of such simple objects as hammers, scissors, airplane silhouettes, and letters. However, most boundary features do not perform well when a small amount of noise is added to the original images (Kauppinen et al. 1995). As mentioned earlier, our experimental data sets are not only much noisier, but the plankton are non-rigid, projection-variant, and often in partial occlusion.

To overcome or at least partially alleviate these difficulties, we turned to features based on mathematical morphology (Matheron 1975; Serra 1982) known as granulometries. Granulometries were introduced in the sixties by Matheron as tools to extract size distributions from binary images (Matheron 1975). The approach is to perform a series of morphological openings of increasing kernel size and to map each kernel size to the number of image pixels being removed during the opening operation at this kernel size. The resulting curve, often called the pattern spectrum (Maragos 1989), maps each size to a measure of the image parts with this size. Typically, the peak of this curve provides the dominant object size in the image. For examples of the application of granulometries, refer to (Vincent 1994a, 1994b).

Beyond pure size information, granulometries actually provide a “pattern signature” of the image to which they are applied and can be used successfully as elements of a feature vector for shape classification problems (Schmitt and Mattioli 1991). Furthermore, the concept of granulometries can easily be extended to gray-scale images. In this context, granulometric curves capture information on object texture as well as shape. Additionally, various types of gray-scale granulometric curves can be computed, depending on the underlying family on openings or closing used. For example, curves based on openings with line segments capture information on bright linear image features, whereas curves based on closings with disk-shaped elements capture information on dark, “blobby” image parts.

To capture information both on linearly elongated and “blobby” image parts, whether dark or bright, we use four types of gray-scale granulometries. These curves are normalized by the total image volume removable by opening or closing, so that they are invariant with respect to illumination and scale changes. For example, an image of the copepod *Oithona* gives rise to a corresponding pattern spectrum characterizing bright, blobby image parts in Figure 2. The peak of this curve characterizes the size and contrast of the body of the organism. A similar curve results for an organism known as a pteropod (a planktonic snail) in Figure 3. The corresponding granulometric curve however, is distinctly different from the previous one, reflecting the more complex body configuration of the pteropod. Several novel gray-scale granulometry algorithms are used to compute the granulometric curves. These

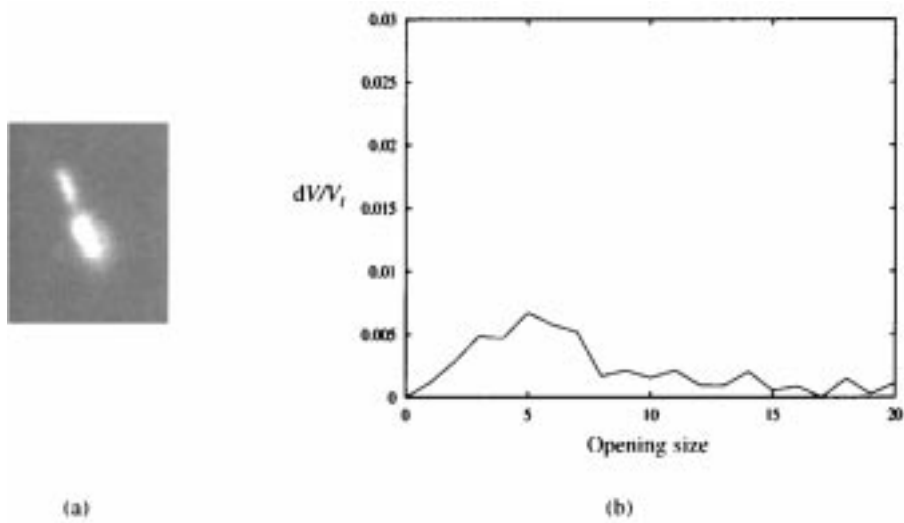


Figure 2. Copepod *Oithona* image (a) and corresponding granulometric curve (b), where dV is the decrease of image “volume”, i.e. the sum of pixel values, from one opening to the next, and V_t is the total volume removable by opening.

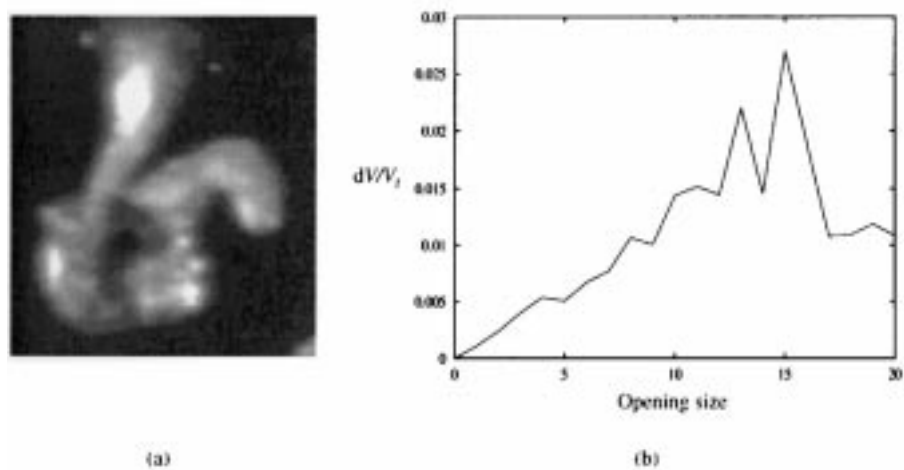


Figure 3. Pteropod image (a) and corresponding granulometric curve (b).

algorithms are orders of magnitude faster than traditional techniques (Vincent 1994b), making possible for real-time application.

2.1.4. Feature selection

Very large feature vectors usually contain much redundant information. Processing requires significant time or computational power, and classifi-

cation results often are poor. For these reasons, decorrelation and feature selection steps are needed. The widely used Karhunen-Loeve transform (KLT) is an ideal feature reduction and selection procedure for our system. Its decorrelation ability serves to decorrelate neighborhood features, and its energy packing property serves to compact useful information into a few dominant features. For a large feature vector, however, the computation of the eigenvectors of the covariance matrix can be very expensive. The dominant eigenvector estimation method, described in Tang (1996) is used to overcome this problem.

As optimal representation features, KLT selected features may not be the best for classification. Additional feature class separability measures are needed to select KLT decorrelated features. We use the Bhattacharyya distance measure in this study. An analytical form of this measure is (Fukunaga 1972)

$$\beta = \frac{1}{8}(\mu_1 - \mu_2)^T \left(\frac{\sigma_1 + \sigma_2}{2} \right)^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \ln \frac{\left| \frac{1}{2}(\sigma_1 + \sigma_2) \right|}{|\sigma_1|^{1/2} |\sigma_2|^{1/2}}, \quad (9)$$

where σ_i and μ_i are the class variance and mean. From (9), we see that β is proportional to both the distance of class means and the difference of class variances. The feature selection criterion is to retain only those decorrelated features with large β value.

2.2. Classification algorithm

We use the learning vector quantization classifier (Kohonen 1987, 1990) for feature classification because it makes weaker assumptions about the shapes of underlying feature vector distributions than traditional statistical classifiers. For this reason, the LVQ classifier can be more robust when distributions are generated by nonlinear processes and are strongly non-Gaussian, similar to the case for our data.

A vector quantization process should optimally allocate M codebook reference vectors, $\omega_i \in R^n$, to the space of n -dimensional feature vectors, $x \in R^n$, so the local point density of the ω_i can be used to approximate the probability density function $p(x)$ (Kohonen 1987). Consequently, the feature vector space is quantized into many subspaces around ω_i , the density of which is high in those areas where feature vectors are more likely to appear and coarse in those areas where feature vectors are scarce.

To use such a vector quantization process in a supervised pattern classification application, Kohonen (1987, 1990) developed the Learning Vector Quantization classifier. First, the codebook reference vectors ω_i s are initialized by either M random feature vector samples or the mean value of the feature

vectors. They are then assigned to a fixed number of known application-specific classes. The relative number of codebook vectors assigned to each class must comply with the a priori probabilities of the classes. A training algorithm is then used to optimize the codebook vectors. Let the input training vector x belong to class C_t , and its closest codebook vector ω_r be labeled as class C_s . The codebook vector ω_i is updated by the learning rules (Kohonen 1987),

$$\begin{aligned} \delta\omega_r &= \alpha(x - \omega_r) & \text{if } C_s = C_t \\ \delta\omega_r &= -\alpha(x - \omega_r) & \text{if } C_s \neq C_t, \\ \delta\omega_i &= 0 & \text{for } i \neq r \end{aligned} \quad (10)$$

where α is the learning rate. Only the closest of the vectors ω_i is updated, with the direction of the correction depending on the correctness of the classification. Effectively, these codebook vectors are pulled away from zones where miscalculations occur, i.e., away from the classification boundary region. After training, the nearest neighbor rule is used to classify the input test vector according to the class label of its nearest codebook vector.

A neural network architecture to implement the LVQ is shown in Figure 4. The network consists of two layers, a competitive layer and a linear output layer. The weights connecting all input neurons with the competitive layer neuron i form the codebook vector ω_i . The net input for each competitive layer neuron is the Euclidean distance between the input vector x and the weight vector ω_i . The output of each neuron is 0 except for the “winner” neuron, whose weight vector has the smallest distance to the input vector and whose output is 1.

The second layer transforms the competitive layer’s neuron class into the final output class. As discussed above, the competitive layer neurons, i.e., the codebook vectors, are assigned to the output classes according to a priori probabilities of the classes. For the example shown in Figure 4, the first three neurons are assigned to Class 1, the next two to Class 2, and the final two to Class 3. Only the non-zero weight connections between the competitive layer and the linear layer are shown in the figure. If any neuron in a particular class wins, the corresponding neuron class in the linear output layer will have an output of 1.

A drawback of the LVQ algorithm is the time consuming training process. To improve it, we use a statistical initial condition and a parallel training strategy introduced by Tang (1996). First, we initialize the neuron weight vectors in the competitive layer using the means and variances of the training vectors in each class. With the initial weight vectors in each class computed by adding random vectors of the class variance to the mean vector of the same class, the training process starts from a good statistical mapping position.

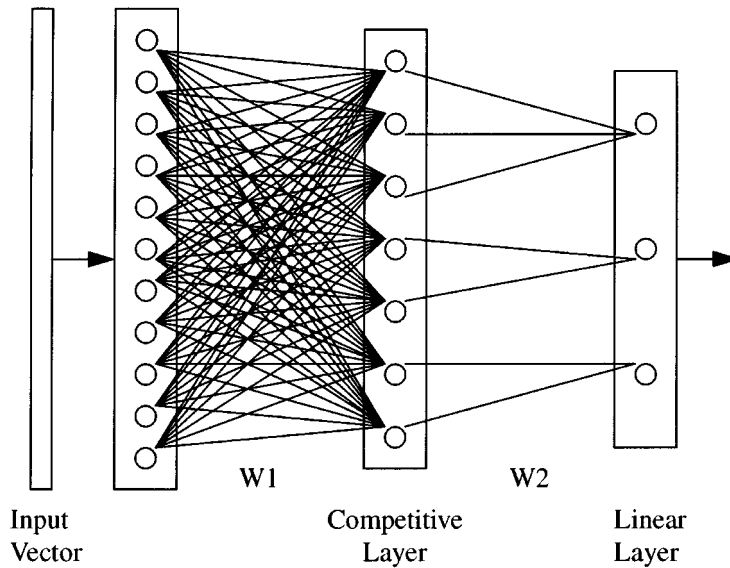


Figure 4. Learning vector quantization neural network.

We make a second improvement by invoking the parallel training strategy (Tang 1996). Traditionally, training samples are randomly selected to be fed into the network one at a time. Therefore only one neuron is updated at each training epoch. The process is quite slow, when there are many neurons to update. The fact that the order in which samples are presented to the network is not important suggests the idea of presenting several training samples in parallel. The only difference this may create is that there may be more than one training sample very near the same neuron. In such a case we select only one of these few samples to update the neuron. This guarantees that parallel processing generates only one update operation on each winning neuron, with results similar to a serial training strategy. The parallel processing can be implemented by either hardware matrix operations or multiprocessor technology.

3. Data Acquisition and Processing

Data acquisition and processing for the current implementation are carried out in two phases. First, a real-time hardware/software system detects in-focus objects, defines a subimage around each detected object, then saves the subimages to disk. Digital storage requirements are reduced by more than two orders of magnitude and the time required to manually identify training images is

accelerated by a similar factor, when compared with manually jogging a videotape editing machine to search for organisms (Davis et al. 1996). In a second phase, more computationally expensive algorithms are applied to the condensed data set for segmentation, feature extraction, and classification.

3.1. *Real-time data acquisition and focused object detection*

The VPR uses a video camera with telephoto lens and a red strobe to obtain magnified images of plankton. The strobe is synchronized with the camera at 60 fields per second. Together, the camera's high resolution (570×485 pixels) and the strobe's short pulse duration allow detailed imaging of the plankton ($10\text{-}\mu\text{m}$ resolution for the 0.5-cm field of view) (Davis et al. 1992). The goal of the VPR's real-time video processing system is to archive to tape digital images of sufficiently large, bright, and in-focus objects as they appear in the video stream.

Live or recorded video and time-code data are sent to the video processing system, which consists of a Sun SPARCstation 20/72 connected to an Imaging Technologies 151 pipelined image processor and a Horita time-code reader (Davis et al. 1996). The image processor can perform real-time (60 field per second) digitization, 3×3 convolutions, rank value filtering, and frame buffer data exchanges with the host workstation. A multi-threaded algorithm on the host is used to supervise the image processor, collect time-code data, compute edge strength, and transfer in-focus subimages to disk.

A simple but effective algorithm has been devised to detect and record objects in real-time (Davis et al. 1996). First, bright large blobs are located in a median filtered image by finding the connected components from run-length lists computed in hardware. For each large blob, the first derivative of the Sobel edge intensity (basically a second derivative of intensity) along the blobs' perimeter is used to reject objects that are out of focus. A gray-scale subimage surrounding any in-focus targets is immediately passed to the workstation for archival and taxonomic classification.

The three main parameters of these algorithms are image intensity threshold, edge strength threshold, and minimum object area in pixels. These are adjusted empirically before data collection based on various factors including lens magnification, transmissivity of the water, and lighting characteristics. Our objective is to achieve a very high probability of detection with a reasonable probability of false alarm, since more intensive processing can be applied to the much smaller data set produced by these algorithms.

On average, about 1 out of 20–60 video fields contains an in-focus object, and only a subimage surrounding the object is saved to disk as an individual file. These object files are time-stamped using the corresponding video time code for precise correlation with ancillary hydrographic and position data

(Davis et al. 1992). This culling process typically reduces the amount of image data to be stored and classified by a factor of 100 or more, thus making the remaining processing computationally feasible (Davis et al. 1996).

3.2. *Data description and processing*

In the following experiment, six classes obtained from nearly 2,000 plankton subimages captured by the VPR are used to test our pattern classification algorithms. They include 133 Calanus, 269 Diat-centr, 658 Diat-chaet, 126 Diatom, 641 Diatomcolo, and 42 Pteropod images. Some sample images for each of the six plankton taxa are illustrated in Figure 1. Half of the images are used as training data and half for testing.

Each gray-scale image is first segmented into a binary image using a simple mean shift method. We use the mean value of the image to threshold the image, then the mean values of the object and the background are computed. The average of the two mean values is used as a new threshold to segment the image again. The process iterates until a stable threshold is reached. Since our images are mostly bimodal, only two iterations give a very good segmentation result, as shown in Figure 5(b). We are currently developing a more robust connectivity-based thresholding technique and will compare the two methods in a future work.

Next, the largest binary object is used to compute the boundary descriptors. This binary image is also used to mask the original gray-scale image to compute moment features. Results for each of these processing steps are illustrated in Figure 5. Granulometric features are computed directly from the original gray-scale images.

4. **Classification Experiments**

A set of experiments was conducted to study the performance of the three types of feature vectors, their combinations, and the improved LVQ classifier. Since there seems to be no simple solution for determining the best network configuration, no exhaustive search was conducted to determine the best network parameters. However, we investigated several network configurations and selected one that appears to be most appropriate for our application. Throughout the experiment, we use the following parameters: 200 competitive layer neurons, learning rate 0.1, parallel training sample number 120 per epoch.

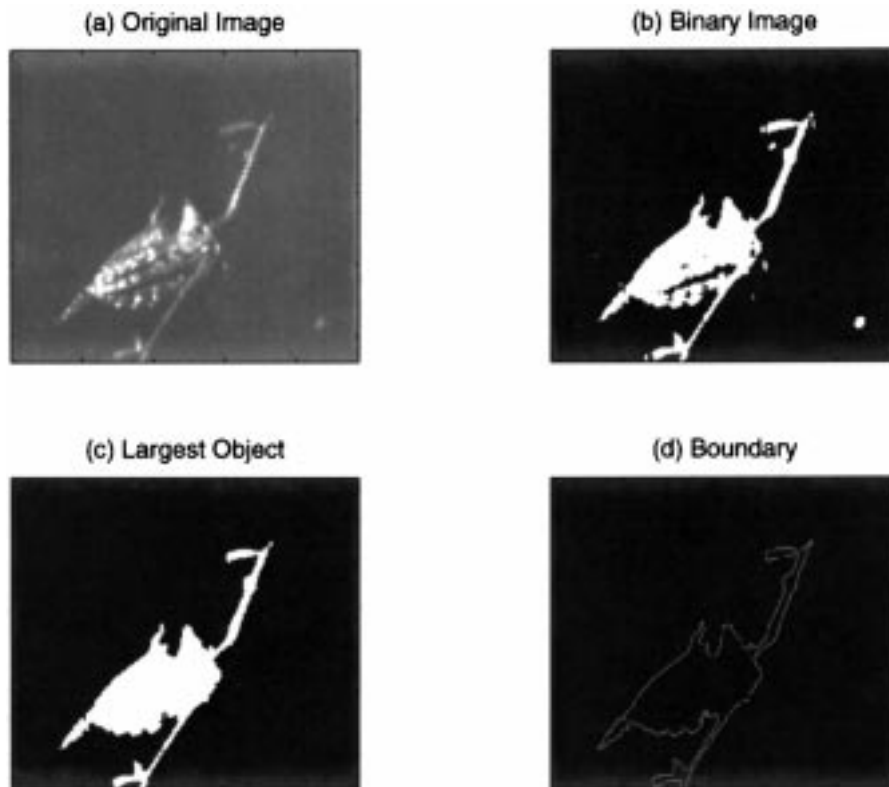


Figure 5. Illustration of the intermediate results of the image processing steps.

4.1. Classification results using individual feature vectors

We first investigate the classification ability of the three individual feature vectors: moment invariants, Fourier descriptors, and granulometries. Comparison of the classification results are given in Table 1. Only 65% classification accuracy is achieved using moment features. This may be attributed in part to the relatively short feature vector length of 7. The images of the same class are also very different in shape because of variations in projection direction, organism body motion, and image occlusions. This shape inhomogeneity also affects the performance of the FD features. For the testing data, only 69% classification accuracy is achieved by the contour FDs. The radius FD features outperform the contour FD features by approximately 10%. This contrasts with Kauppinen et al. (1995), where contour FDs give better results than the radius FDs. Such a discrepancy in results may be caused by the differences between the two data sets. More conclusive comparison studies need to be done before we draw a firm conclusion. The sampling number we used for

Table 1. Classification rates (%) on six classes of plankton images using individual feature vectors: Moment invariants, Fourier descriptors (FD), and Granulometries

Feature types	Original feature length	Number of selected features	Correct classification rate (%)		
			Training	Testing	All data
Moment invariants	7	7	67.74	63.13	65.44
Contour FD	360	28	83.7	69.2	76.5
Radius FD	180	21	94.6	78.0	86.3
Granulometries	160	29	97.8	86.4	92.1

the boundary function is 360 points. This number is much higher than for many previous studies, because plankton have noisier, more irregular boundaries requiring a higher sampling rate to capture high-frequency information. The granulometry features give the best performance with a better than 90% accuracy. This demonstrates the features' insensitivity to occlusion, image projection direction, and body motion because of rich three-dimensional texture and shape information captured.

4.2. Classification results using combined feature vectors

The confusion matrices of the classification results using the three individual features are given in Tables 2–4, where the columns are true class labels, the rows are the resulting labels. Notice that the shapes of these matrices are quite different. The moments are good at distinguishing Diat-centr and Diatom; the radius FD generates good results on Diat-centr and Diat-chaet; the granulometry features perform well on all classes, except on the Diat-centr. All these suggest that the discrimination abilities of the three feature types may be distinct when recognizing different object characteristics. To form a more complete description of the plankton patterns, we combine all three feature vectors into a single feature vector. Results are shown in Table 5. The combined vector yields 95% classification accuracy, which is comparable with what a trained biologist can achieve by using conventional manual techniques (Davis 1982). Notice also that the feature length is condensed to around 20 from several hundreds, demonstrating the efficiency of our feature selection approach.

Not all combinations of feature vectors in Table 5 show an improvement in results. For example the combined contour and radius FDs perform worse than the radius FD features alone. Adding moment features to the granulometry feature vector only improves results slightly. The short moment feature vector seems to contain only a subset of information contained in the granulometry features.

Table 2. Confusion matrix of the moment invariant features

Names	Calanus	Diat-centr	Diat-chaet	Diatom	Diatomcolo	Pteropod
Calanus	41	0	71	0	20	4
Diat-centr	3	241	3	0	65	8
Diat-chaet	67	3	446	8	180	11
Diatom	0	0	10	116	0	0
Diatomcolo	21	23	127	2	373	13
Pteropod	1	2	1	0	3	6

Table 3. Confusion matrix of the radius Fourier descriptors

Names	Calanus	Diat-centr	Diat-chaet	Diatom	Diatomcolo	Pteropod
Calanus	71	0	9	3	26	4
Diat-centr	0	262	2	0	4	1
Diat-chaet	22	6	602	5	66	3
Diatom	2	0	3	117	1	3
Diatomcolo	34	1	42	1	539	9
Pteropod	4	0	0	0	5	22

Table 4. Confusion matrix of the granulometry features

Names	Calanus	Diat-centr	Diat-chaet	Diatom	Diatomcolo	Pteropod
Calanus	117	2	15	2	7	0
Diat-centr	0	239	7	0	0	0
Diat-chaet	15	27	600	5	25	0
Diatom	1	0	16	116	2	0
Diatomcolo	0	1	20	3	607	0
Pteropod	0	0	0	0	0	42

Table 5. Classification rates (%) on the six classes of plankton images using combined feature vectors

Feature types	Original feature length	Number of selected features	Correct classification rate (%)		
			Training	Testing	All data
Contour FD & Radius FD	540	29	90.4	76.8	83.6
Moments & granulometry	167	29	97.5	87.5	92.5
Granulometry & radius FD	340	24	98.7	91.5	95.1
Moments & granulometry & radius FD	347	19	98.0	92.2	95.1

Table 6. Confusion matrix of the combined moments, radius FDs, and granulometry features

Names	Calanus	Diat-centr	Diat-chaet	Diatom	Diatomcolo	Pteropod
Calanus	120	0	9	1	8	1
Diat-centr	0	262	7	0	1	2
Diat-chaet	13	6	627	2	23	0
Diatom	0	0	5	120	0	0
Diatomcolo	0	1	10	3	609	0
Pteropod	0	0	0	0	0	39

The combined feature vector confusion matrix is shown in Table 6. Although the DIAT-CHAET images have a classification accuracy of 95.3%, about the average of all classes, many other class images are misclassified as DIAT-CHAET, such as all misclassified CALANUS, most misclassified DIAT-CENTR and DIATOMCOLO images. This is probably because the versatile DIAT-CHAET images have texture and shape structures similar to those of other class images. For example, the body texture of some DIAT-CHAET look quite similar to that of CALANUS, and some small samples of DIAT-CHAET may be confused with DIAT-CENTR. All images are sorted by a trained biologist and sometimes only very subtle characteristics are used to judge the occluded plankton images. Some human errors in the identification were discovered using the automated method. Given the data quality, the overall classification rate is very encouraging.

4.3. Comparison of LVQ training methods

A traditional serial training method is compared to the new parallel algorithm in Figure 6. The three lines show the progression of training data classification accuracy with increasing number of training epochs using three methods: the traditional training method with all neurons initialized with the mean value of the training samples, the traditional training method with the statistical initial condition, and the new parallel training method. The new method reaches 98% training accuracy within 100 epochs, while the traditional methods achieve 95% accuracy using nearly 10000 epochs.

To further compare the traditional method with the new parallel training method, we use scatter plots of the first two dimensions of the training sample feature vectors and their corresponding neuron weight vectors at several training stages in Figure 7. As the training progresses, the neuron weight vectors gradually start to match the topology of the feature vector space. Comparing Figures 7(a) and (b), we see that the new method apparently maps the density of the feature vector space better and more quickly.

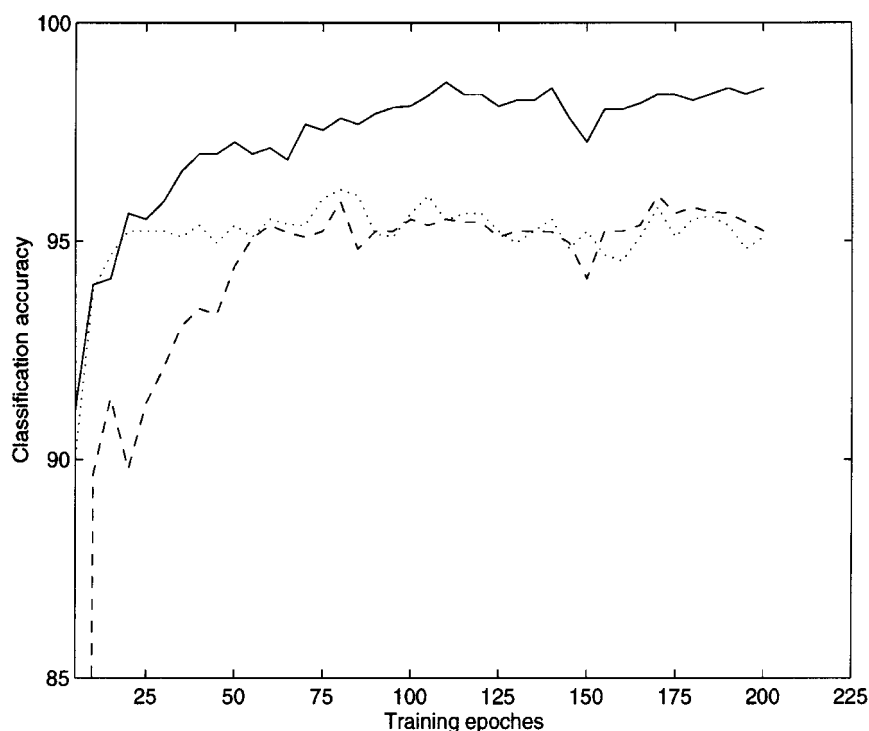
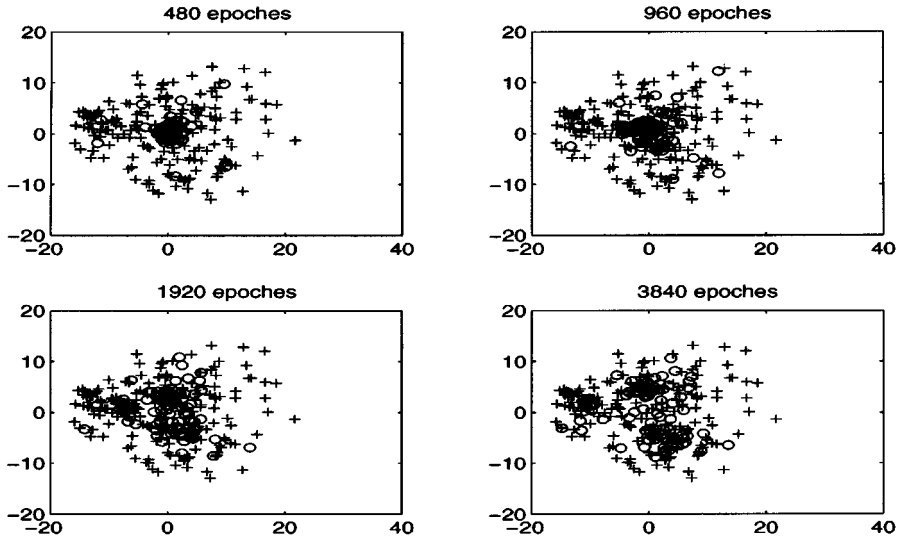


Figure 6. Comparison of the three training methods. The dashed line represents the classification accuracy progression with training epochs using the traditional training method and the mean value initial condition. The dotted line is for the traditional training method with statistical initial condition. The solid line is the result of the new parallel training method. The epoch numbers on the horizontal axis for the two traditional training methods should be multiplied by 120.

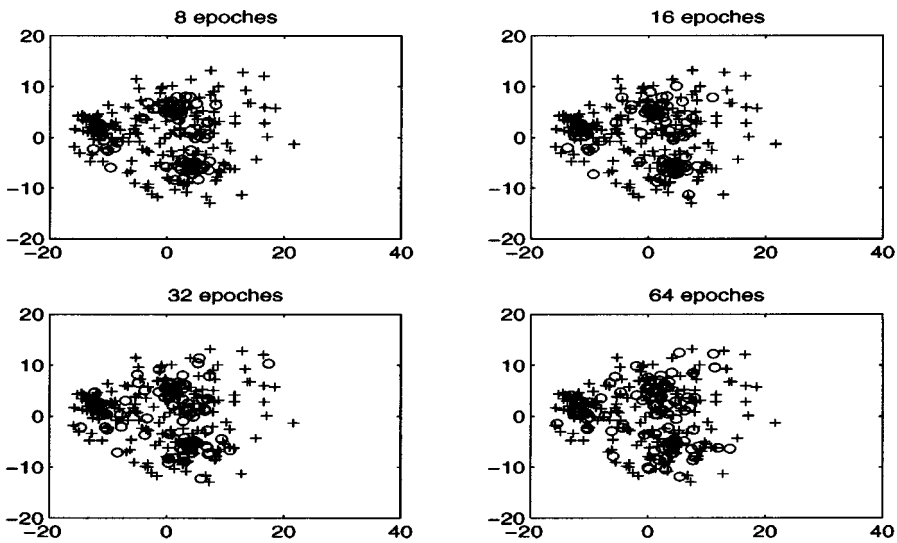
5. Conclusions and Future Work

Our experimental results clearly demonstrate that the combined feature vector is better than any single feature type. Of the individual feature types, the granulometry vector contains more information than conventional shape descriptors. The KLT and Bhattacharyya feature selection method successfully decorrelates and compacts a large feature vector into a small description vector. Together with the improved LVQ classifier, 95% classification accuracy is achieved.

With the VPR system, including the data acquisition unit and the pattern classification system, real-time automatic sorting of plankton into taxonomic categories becomes possible for the first time. This ability will allow rapid acquisition of size-dependent taxonomic data over a wide range of scales



(a) Training using traditional method.



(b) Parallel training with statistical initial condition.

Figure 7. Comparison of the traditional and the new parallel training methods using scatter plots of the first two feature dimensions and their corresponding neuron weight vectors at several training stages. The '+' markers represent the six classes of training samples and the 'o' markers represent the hidden layer neurons.

in a dynamic oceanographic environment, providing new insights into the biological and physical processes controlling plankton populations in the sea (Davis et al. 1992).

To achieve this goal, though, much work remains. First, biologists are interested in classifying many more plankton species than tested here, many of which are quite similar. To maintain a high degree of classification accuracy, we must develop more distinct pattern features. A hierarchical classification system may also be necessary to classify major species and subspecies in several steps. We also intend to integrate our classification algorithms with the current VPR processing system to carry out real-time taxonomic identification at sea.

Acknowledgements

We thank P. H. Wiebe for his comments. This work is supported by the Office of Naval Research through grant N00014-93-1-0602. This is contribution number 9214 of the Woods Hole Oceanographic Institution.

References

- Davis, C. S. (1982). *Processes Controlling Zooplankton Abundance on Georges Bank*. Ph.D. Thesis, Boston University.
- Davis, C. S., Gallager, S. M., Berman, N. S., Haury, L. R. & Strickler, J. R. (1992). The Video Plankton Recorder (VPR): Design and Initial Results. *Arch. Hydrobiol. Beih.* **36**: 67–81.
- Davis, C. S., Gallager, S. M. & Solow, A. R. (1992). Microaggregations of Oceanic Plankton Observed by Towed Video Microscopy. *Science* **257** (10 July): 230–232.
- Davis, C. S., Gallager, S. M., Marra, M. & Stewart, W. K. (1996). Rapid Visualization of Plankton Abundance and Taxonomic Composition Using the Video Plankton Recorder. *Deep-Sea Research II*, Vol. 43, No. 7–8, pp. 1947–1970.
- Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*. Academic Press: New York.
- Gonzalez, R. C. & Wintz, P. (1987). *Digital Image Processing*. Addison-Wesley.
- Hu, M. K. (1962). Visual Pattern Recognition by Moment Invariants. *IRE Trans. Information Theory* **IT-8** (Feb.): 179–187.
- Kauppinen, H., Seppanen, T. & Pietikainen, M. (1995). An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification. *IEEE Trans. on Pattern Analysis and Mach. Intell.* **17**(2) (Feb.): 201–207.
- Kohonen, T. (1987). *Self-Organization and Associative Memory*, 2nd Edition. Springer-Verlag: Berlin.
- Kohonen, T. (1990). The Self-Organizing Map. *Proceedings of the IEEE* **78**(9) (Sept.): 1464–1480.
- Maragos, P. (1989). Pattern Spectrum and Multiscale Shape Representation. *IEEE Trans. on Pattern Analysis and Mach. Intell.* **11**(7) (July): 701–716.
- Matheron, G. (1975). *Random Sets and Integral Geometry*. John Wiley and Sons: New York.
- Owen, R. W. (1989). Microscale and Finescale Variations of Small Plankton in Coastal and Pelagic Environments. *Journal of Marine Research* **47**: 197–240.

- Paffenhofer, G. A., Stewart, T. B., Youngbluth, M. J. & Bailey, T. G. (1991). High-Resolution Vertical Profiles of Pelagic Tunicates. *Journal of Plankton Research* **13**(5): 971–981.
- Persoon, E. & Fu, K. S. (1977). Shape Discrimination Using Fourier Descriptors. *IEEE Trans. on Sys., Man, and Cyber* **7**(3) (Mar.): 170–179.
- Reeves, A. P., Prokop, R. J., Andrews, S. E. & Kuhl, F. P. (1988). Three-Dimensional Shape Analysis Using Moments and Fourier Descriptors. *IEEE Trans. on Pattern Analysis and Mach. Intell.* **10**(6) (Nov.): 937–943.
- Reiss, T. H. (1991). The Revised Fundamental Theorem of Moment Invariants. *IEEE Trans. on Pattern Analysis and Mach. Intell.* **13**(8) (Aug.): 830–834.
- Reti, T. & Czinege, I. (1989). Shape Characterization of Particles via Generalized Fourier Analysis. *Journal of Microscopy* **156**(Pt. 1) (Oct.): 15–32.
- Schmitt, M. & Mattioli, J. (1991). Shape Recognition Combining Mathematical Morphology and Neural Networks. *SPIE: Application of Artificial Neural Network*. Orlando, Florida (April).
- Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press: London.
- Tang, X. (1996). *Transform Texture Classification*. Ph.D. Thesis, Massachusetts Institute of Technology, the MIT/Woods Hole Oceanographic Institution Joint Program.
- Teh, C. H. & Chin, R. T. (1991). On Image Analysis by the Methods of Moments. *IEEE Trans. on Pattern Analysis and Mach. Intell.* **10**(4) (July): 496–513.
- Vincent, L. (1994a). Fast Opening Functions and Morphological Granulometries. *Proc. Image Algebra and Morphological Image Processing*, SPIE 2300, 253–267. San Diego (July).
- Vincent, L. (1994b). Fast Grayscale Granulometry Algorithms. *EURASIP Workshop ISMM '94, Mathematical Morphology and Its Applications to Image Processing*, 265–272. France (Sept.).
- Zahn, C. & Roskies, R. Z. (1972). Fourier Descriptors for Plane Closed Curves. *IEEE Trans. on Computers* **21**(3) (Mar.): 269–281.