Accurate automatic quantification of taxa-specific plankton abundance using dual classification with correction

Qiao Hu[1,*], Cabell Davis[2]

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

ABSTRACT: Optical imaging samplers are becoming widely used in plankton ecology, but image analysis methods have lagged behind image acquisition rates. Automated methods for analysis and recognition of plankton images have been developed, which are capable of real-time processing of incoming image data into major taxonomic groups. The limited accuracy of these methods can require significant manual post-processing to correct the automatically generated results, in order to obtain accurate estimates of plankton abundance patterns. We present here a dual-classification method in which each plankton image is first identified using a shaped-based feature set and a neural network classifier, and then a second time using a texture-based feature set and a support vector machine classifier. The plankton image is considered to belong to a given taxon only if the 2 identifications agree; otherwise it is labeled as unknown. This dual-classification method greatly reduces the false positive rate, and thus gives better abundance estimation in regions of low relative abundance. A confusion matrix is computed from a set of training images in order to determine the detection and false positives rates. These rates are used to correct abundances estimated from the automatic classification results. Aside from the manual sorting required to generate the initial training set of images, this dual-classification method is fully automatic and does not require subsequent manual correction of automatically sorted images. The resulting abundances agree closely with those obtained using manually sorted results. A set of images from a Video Plankton Recorder was used to evaluate this method and compare it with previously reported single-classifier results for major taxa.

KEY WORDS: Plankton · Video · Sampling · Pattern recognition · Real-time · Rejection

INTRODUCTION

Processes controlling distributions of marine plankton operate across a wide range of time–space scales, and predicting abundance patterns in these organisms remains an elusive goal, despite more than 100 yr of research (Marine Zooplankton Colloquium 1989, 2001). Due to the inherently patchy nature of plankton (Fasham 1978, Haury et al. 1978, Omori & Ikeda 1984, Mackas et al. 1985), traditional sampling methods have not been able to adequately resolve their distributions. New optical imaging systems have the capability to sample major plankton taxa, including delicate forms and particulate matter, nearly continuously over a broad range of time and space scales (Davis et al. 1992b, Gallager et al. 1996, Norrbin et al. 1996, Ashjian et al. 2001, 2005, in press). Due to the large amount of image data from optical samplers (e.g. Video Plankton Recorder, VPR; Davis et al. 1992a, 1996, 2005), fully automatic analysis is highly desirable.

The VPR was designed to identify plankton according to major taxonomic group (e.g. copepods, chaetognaths; Davis et al. 1992a,b). VPR images are collected

at ship speeds of 4 to 6 m s^{-1} and at a working distance[1] of 30 to 50 cm, so that the optical quality of the images, although good, is obviously less optimal than that obtainable under a laboratory microscope. Existing optical imaging samplers yield an intermediate level of taxonomic resolution, providing information beyond the biomass and size distributions obtained from acoustical samplers (Holliday et al. 2003, Wiebe et al. 2002) and particle size/shape counters (Herman 1992, Herman et al. 2004). Automatic identification of field-collected plankton images, even to the major taxa, has proven difficult, owing to lighting variability, projection variance, and occlusion (cropping).

Automatic identification involves extracting a set of n features from an image, mapping these features to a point in the n-dimensional feature space, and then dividing the feature space into regions belonging to different taxa. The decision boundary is based on the locations of training images (labeled images) in the feature space. Thus, automatic identification requires manually sorting a subset of images that are used to 'build' or 'train' the classifier. The training process involves finding the optimal decision boundary in the feature space.

Research on automatic plankton classification has been on-going for more than 2 decades (Jeffries et al. 1980, 1984, Rolke & Lenz 1984, Gorsky et al. 1989, Berman et al. 1990, Hofstraat et al. 1994, Costas et al. 1995). Early studies focused on well-controlled laboratory-obtained images. More recently, artificial neural networks (NN) and support vector machines (SVM) have been introduced to this field (Simpson et al. 1991, Boddy & Morris 1993, Culverhouse et al. 1994, 2003 Tang 1998, Tang et al. 1998, Toth & Culverhouse 1999, Davis et al. 2004, Luo et al. 2004, Hu & Davis 2005). Most of the previous classification systems were trained and tested on a small set of selected distinctive images. Davis et al. (2004) applied real-time automatic classification methods on field data to estimate mean abundance of major taxonomic groups while sampling underway at sea. They showed that including difficult and human unidentifiable images in the training or testing of a classifier greatly reduces the prediction accuracy (cf. Luo et al. 2004). On the other hand, their work also showed that it was important to include an 'other' category (including rare and unidentifiable taxa) to better estimate abundance (cf. Grosjean et al. 2004).

Early identification systems mainly used shape-based features (e.g. morphological measurements and Fourier descriptors). Due to the early success on laboratory-obtained images, shape-based features are still a mainstay in the field of plankton recognition. Recently, Hu & Davis (2005) suggested using texture-based features to help overcome difficulties in classification of field-collected images, such as occlusion,

non-linear illumination, and projection variance. The automatic classification system used in these previous studies (Davis et al. 2004, Hu & Davis 2005) yielded abundance patterns that were, in general, very similar to those obtained by manually sorting all the images in the tow. However, in the regions of low relative abundance, the abundance estimation from automatic classification was often overestimated, due to the relatively high proportion of false alarms (false positives). A manual correction step (i.e. manual removal of false positives) was needed to obtain accurate abundance in these regions.

In the present paper, we describe a new dual-classification method in which a rejection metric is obtained by 'voting' with 2 classifiers: (1) an NN classifier built from shape-based features and (2) an SVM classifier built from texture-based features. Both classifiers must agree on the identification of an image for it to be considered true, otherwise it is classified as 'unknown'. Abundance estimation from the dual-classification system was corrected based on detection and false-alarm rates. After correction, the abundance estimation from the automatic classification system was compared to that derived from manually sorted results and from previous classification results.

MATERIALS AND METHODS

Data description. Our data came from a 24 h VPR tow in the Great South Channel off Cape Cod, Massachusetts, in June 1997. The images were taken by a high-magnification camera, which has an image volume of 0.5 ml. The total sampled volume during the deployment was about 2.6 m^3. There were >20 000 images in this tow, all of which were sorted manually into 7 major categories[2]. The major groups included 5983 copepods, 3656 rod-shaped diatom chains, 478 *Chaetoceros* chains, 2394 *Chaetoceros socialis* colonies, 1796 hydroid medusae, 1650 marine snow, and 4304 'other'. Most of the images in 'other' category were unidentifiable by human experts; the rest were rare species, including coil-shaped diatom chains, ctenophores, chaetognaths, and copepod nauplii (see Davis et al. 2004, Hu & Davis 2005). From previous experience (Davis et al. 2004), we found that 200 images per taxon was a reasonable number for the training set. That limit automatically disqualifies most rare taxa from being used as separate categories in the

---

[1] Working distance here is the distance from the center of the imaged volume to the optical window of the underwater camera housing

[2] Example images of these 7 taxa are given in Davis et al. (2004) and Hu & Davis (2005)

classifier. Very rare taxa, the total numbers of which were between 200 and 400, were also grouped into the 'other' category. The manual identification of each image was treated as the basis for comparison among different automatic classifiers.

Pattern representations/feature measurements. Five different types of features were used in our system, including shape-based features (moment invariants, morphological measurements, Fourier descriptors, and granulometry curves) and texture-based features (co-occurrence matrix).

Moment invariants: Moment invariants, introduced by Hu (1962), are based on normalized central moments, and are translation, rotation, and scale invariant. They have been widely used in plankton identification (Jeffries et al. 1980, 1984, Berman et al. 1990, Tang 1998, Tang et al. 1998, Davis et al. 2004, Luo et al. 2004).

Morphological measurements: Jeffries et al. (1980, 1984) first used 7 morphological measurements as features to identify zooplankton. The concept of using morphological measurements as features in plankton recognition has been commonly accepted ever since then (Berman et al. 1990, Davis et al. 2004, Lou et al. 2004). In the present study, 6 morphological measurements were used as part of the shape-based feature set: (1) a shape factor based on the perimeter and area of the object; (2) a ratio of the maximum and minimum principal moments of the object; (3) a ratio of the longest to shortest dimension of the bounding box surrounding the object; (4) a ratio of the width at the center to the shortest dimension of the bounding box; (5) a ratio of the left $\frac{1}{4}$-width to the shortest dimension of the bounding box; and (6) a ratio of the right $\frac{1}{4}$-width to the shortest dimension of the bounding box (Davis et al. 2004).

Fourier descriptors: Fourier descriptors (FD) are well-studied invariant features used to describe the contour of an object. Depending on what functions are used to describe the contour, FDs can be grouped into 3 major categories, namely tangent-angle FDs (Zahn & Roskies 1972), complex FDs (Granlund 1972), and elliptic FDs (Kuhl & Giardina 1982). FDs are the Fourier coefficients when the contour function is approximated by a Fourier series. Normalized FDs are translation, rotation, and scale invariant. FDs were used as features to classify plankton images (Jeffries et al. 1980, 1984, Berman et al. 1990). In this study, we used a centroidal radius-vector function (distances from the centroid to perimeter pixels) as the contour model[3]. The first 64 elements of the normalized power

spectrum, obtained from the Fourier transform of the centroidal radius-vector function, were used as a feature set (Tang 1998, Tang et al. 1998, Davis et al. 2004).

Granulometry: The concept of granulometry was introduced by Matheron (1975) to study the size distribution of binary images. The operation involves a series of opening/closing operators, with structuring elements of increasing/decreasing size (Serra 1982). Tang et al. (1998) first used granulometry features to classify plankton images. They found the granulometry was more powerful in discriminating plankton images than common moment invariants and Fourier descriptors. However, these operators are computationally expensive. Fast algorithms (Vincent 1993, Meijster & Wilkinson 2002) were developed for very limited structural elements. In the present study, Vincent's (1993) algorithm was used to calculate the linear opening and closing spectra, as well as pseudo-opening and -closing spectra. Each spectrum has 40 elements, resulting in 160 features for granulometry.

Co-occurrence matrix: A gray level co-occurrence matrix (GLCM) was first proposed by Haralick et al. (1973) as a texture feature to classify satellite images. It is based on estimation of the joint probability distribution of pairs of pixels with gray-scale level i and j, spatial distance d, and angle θ in an image. Hu & Davis (2005) first used GLCM to classify plankton images. They found that these texture-based features were more powerful than shape-based features for classifying field-collected plankton images, due to occlusion, non-linear illumination, and 3D to 2D projection variance of the images.

Feature extraction. *Shape-based features:* All the shaped-based features were stacked into 1 feature vector. The features included 7 moment invariants, 6 morphological measurements, 64 Fourier descriptors, and 160 granulometry measurements (Davis et al. 2004). Each feature element was normalized to have zero mean and unit standard deviation. Principal component analysis was applied on the feature vector to calculate dominant eigenvectors. The first 30 features associated with the largest eigenvalues were saved as the feature vector, and the corresponding feature vectors were saved as a transformation matrix (Tang et al. 1998).

Texture-based features: Four different distance (1, 4, 8, 16 pixels) pairs and 4 different angles (0°, 45°, 90°, 135°) were used to generate co-occurrence matrices. For each separation distance, there were 4 co-occurrence matrices from 4 different angles. Only the mean and range of these matrices were used to achieve relative rotation invariance. Normalization was also applied to the resulting matrices to achieve scale invariance. The angular second moment (energy), contrast, correction, variance, inverse-difference mo-

---

[3] Radius-vector functions are only suitable for star-shaped contour models. Most plankton images are not star shaped. However, in a recent experiment, we found the performance difference for different FDs to be very small

ment, and the entropy, sum entropy, and difference entropy of these matrices (Haralick et al. 1973) were calculated and used as feature vector elements. Each feature element was further normalized to have zero mean and unit standard deviation (Hu & Davis 2005). For each image, 64 features were used.

Classifiers. Learning vector quantization neural network and support vector machine classifiers were used in this study.

Learning vector quantization: Learning vector quantization (LVQ) is a supervised version of vector quantization. Its objective is to learn a set of prototypes (codebooks) which best represent each class. We implemented it with an artificial neural network (Tang et al. 1998, Davis et al. 2004). LVQ neural network (LVQ-NN) is a method to divide n-dimensional feature space into different taxonomic regions by fitting neurons to the training data. The neural network has 2 layers, namely a competitive layer and a linear output layer. The complexity of the neural network was based on the number of training images and the number of taxa in the classifier. For the 200 training images per taxon, we used 20 neurons per taxon for the competitive layer. The number of output-layer neurons is equal to the number of taxa. The weights of the neurons for each class were initialized to the mean of the training feature vectors for that class plus a small random value. The network was trained by randomly presenting the training samples one at a time to the network. For a given training sample, the nearest neuron was found (i.e. shortest Euclidian distance in n-space). The taxon assigned to this nearest neuron was the 'predicted' taxon. If the prediction was correct, the weights of this nearest neuron were updated in such a way to move it a step closer to that sample in the feature space. Otherwise, the weights of the nearest neuron were updated such that it was pushed a step away from that sample in the feature space. Over-training was avoided by stopping training early.

Support vector machine: The support vector machine (SVM) is a margin-based linear machine (Vapnik 1995, 1998). Instead of using neurons, the basic idea of SVM is to find a hyperplane that separates the training samples with maximum margin. The capacity of linear SVM is often limited. In order to deal with the nonlinear problem, an intermediate step is taken to map original features to a much higher dimensional space; a hyperplane is then constructed on that higher space. The trick of the SVM is to pick certain mapping functions, so that the mapping is equivalent to applying a kernel function on the original features. SVM is closely related to structural risk minimization and regularization theory. It has shown an excellent generalization property and resistance to over-training in a number of real-world problems (Mohan et al. 2001, Decoste &

Schölkopf 2002, Mitchell et al. 2004, Hu & Davis 2005, Kim et al. 2005). SVM is primarily a binary classifier. Three approaches are often used to extend SVM to the multi-class case, namely a 1-versus-all approach, a pairwise approach, and an error-correcting output codes approach. From our previous study (Hu & Davis 2005), we found that the SVM classifier was not very sensitive to kernel types and kernel parameters. In this study, we chose a linear kernel function to avoid extra labeled validation samples which were needed in kernel parameter selection. The pairwise approach was used, since it yielded balanced training in our case (Hu & Davis 2005).

Dual-classification system. The schematic diagram of the dual-classification system is shown in Fig. 1. During the training phase, 2 classifiers were built in parallel. An LVQ-NN classifier was built from shape-based features as discussed in the 'Feature extraction' section. At the same time, an SVM classifier was built using texture-based features from the same training samples. In the classification phase, shape-based and texture-based features were calculated from all the images. An LVQ-NN classifier made the identification based on shape-based features, while an SVM classifier made the identification based on texture-based features. In the end, a classifier committee was called. If the labels predicted by 2 classifiers belonged to the
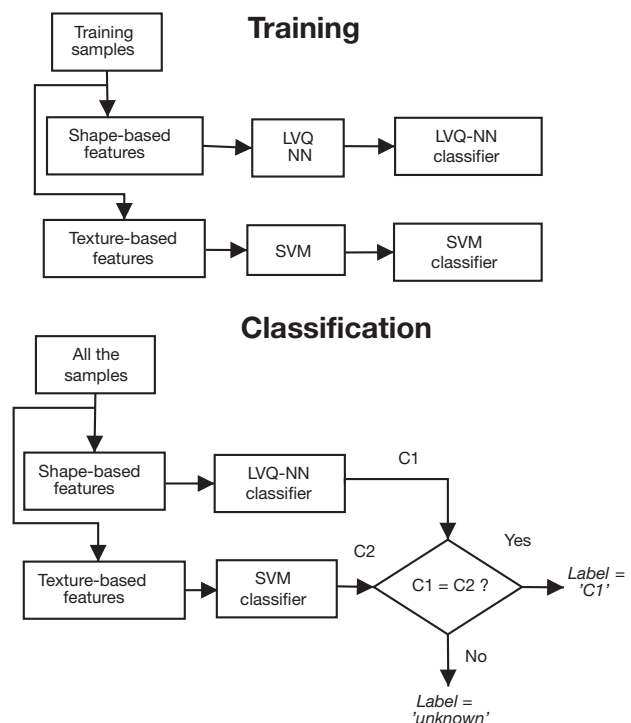

Fig. 1. Schematic diagram of dual-classification system. LVQ: learning vector quantization; NN: neural network; SVM: support vector machine

same class, the sample was labeled as that class. Otherwise, the sample was labeled as 'unknown'.

Classification performance evaluation and abundance correction. *Confusion matrix:* The confusion matrix is used to assess the accuracy of automatic classification. The number of images manually sorted by a human is given in the columns (1 column per taxon), while the number of images automatically classified by the computer is given in the rows (1 row per taxon). Diagonal elements correspond to agreement between human and machine. In the present study, the confusion matrix of the dual-classification system was built in the following way. First, a 7-row by 6-column matrix was built from a set of training images (200 per taxon) for 6 dominant taxa (Table 1) using the leave-one-out method (cross-validation; Davis et al. 2004). The 7th row in this matrix contains the 'unknown' counts. Second, 200 images that had been manually sorted into the category 'other' were classified using the dual-classification system to fill in the 7th column[4]. The resulting 7 × 7 matrix was used as a confusion matrix for the classification system (i.e. Table 1).

From the matrix, some simple indexes of classifier performance can be calculated. The most used indexes are 'probability of detection' (also known as 'sensitivity' or 'probability of true positives') and 'probability of false alarm' (also known as 'probability of false positives'). The probability of detection, P_D, measures the probability that the classifier will identify an image correctly, i.e. P_D = true positive counts/(true positive

counts + false negative counts). The probability of false alarm is the probability that an image will be classified as a given taxon when it does not actually belong to that taxon. Another related concept is specificity, SP (Baldi & Brunak 2001), which is the probability that the classifier's prediction is correct for each taxon, i.e. SP = true positive counts / (true positive counts + false positive counts). Probability of detection and specificity of each taxon were calculated from the confusion matrix to correct the abundance estimation.

Abundance correction: If P_D and SP of a classification system for each taxon are always the same, plankton abundance estimated from the classification system will be perfect although the classification system itself is not perfect. In reality, P_D and SP for each taxon may change for different-sized evaluation data sets. In particular, the SP of a taxon is positively related to the relative abundance of the taxon. However, if the variation in P_D and SP of a classification system for each taxon is relatively small in the study area, we can automatically correct the abundance estimation of a classification system using the following steps: (1) estimate P_D and SP for each taxon from the confusion matrix; and (2) scale the abundance estimation from the classification system for each taxon by the ratio SP/P_D for that taxon. In Davis et al. (2004), false positives were removed manually, so that that the specificity of each taxon was unity, and the correction factor for each taxon was 1/P_D. This correction method is different from the statistical correction method proposed by Solow et al. (2001).

RESULTS AND DISCUSSION

The first 25 images classified as copepods and *Chaetoceros socialis* colonies by the dual-classification system and by the single neural network classifier used previously (Davis et al. 2004) are shown in Fig. 2. For taxa with high relative abundance (i.e. copepods), the performance of dual and single classifiers is very similar, which implies the 2 classification systems have very close probability of detection. On the other hand, for taxa with lower relative abundance (i.e. *C. socialis*), the dual-classification system performs much better, because it has a far lower false alarm rate (Fig. 2). The dual-classification system has much higher specificity for *C. socialis* in regions of low relative abundance (cf. Tables 1 & 2). In other words, the dual-classification system makes the specificity less variable with changes in relative abundance of a taxon, which makes automatic correction of classification results possible (Tables 1 & 2).

Table 1. Confusion matrix of the dual-classification system, using leave-one-out method. Randomly selected images (200 per category) from EN302 VPR Tow 7 were used to construct the confusion matrix. C1: copepods; C2: rod-shaped diatom chains; C3: *Chaetoceros* chains; C4: *Chaetoceros socialis* colonies; C5: hydroid medusae; C6: marine snow; C7: other; C7*: unknown; P_D: probability of detection (%); SP: specificity (%); NA: not applicable. True counts (i.e. human counts) for a given taxa are given in the columns, while counts by the classification system are given in the rows. Correct identifications by the computer are given along the main diagonal, while the off-diagonal entries are incorrect identifications by the computer. All data are counts, except in the last row and last column, which are percent values. Although images of the category 'other' are not needed to train the classification system, they are necessary to evaluate it

	C1	C2	C3	C4	C5	C6	C7	SP
C1	146	0	3	2	1	2	59	69
C2	2	176	2	0	1	1	13	90
C3	0	0	122	1	3	1	2	95
C4	0	0	0	145	2	8	12	87
C5	0	0	5	0	111	6	4	88
C6	1	0	0	3	4	98	4	89
C7*	51	24	68	49	78	84	106	23
P_D	73	88	61	73	56	49	53	NA

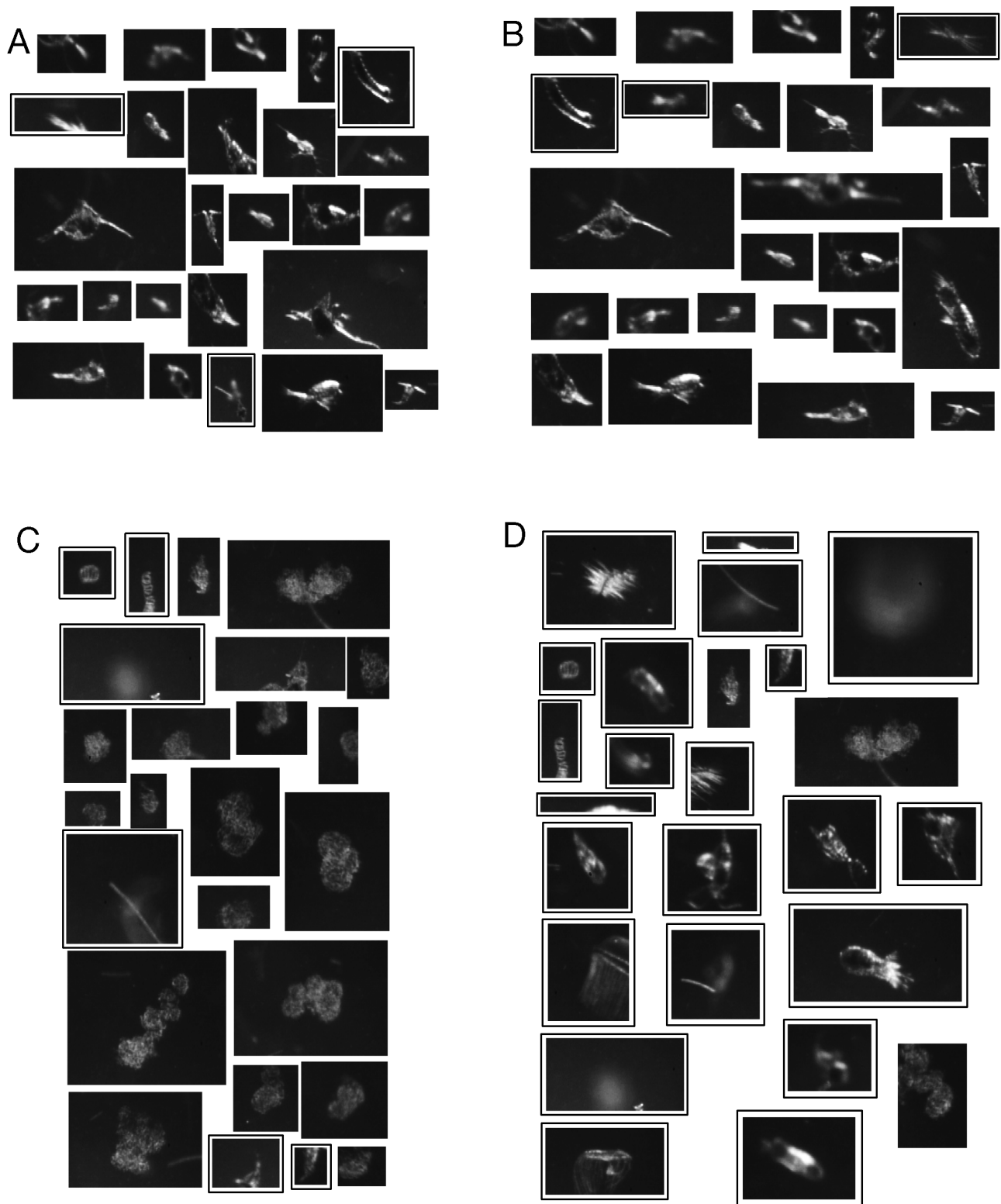[4] The last classifier built in the leave-one-out method was used

Fig. 2. Automatically classified images: comparison of results for (A,C) dual-classification system and (B,D) single neural-network classifier. The first 25 images classified as (A,B) copepods and (C,D) *Chaetoceros socialis* by the dual-classification system and LVQ-NN classifier are shown. For taxa having relatively high abundance, such as copepods, both systems yield very similar results (21 out of 25 were the same). In contrast, for taxa having relatively low abundance, such as low-abundance regions of *C. socialis*, the dual-classification system has much higher specificity (fewer false alarms). False positives are denoted with black frames

Abundance estimates of 6 dominant taxa were compared according to the following methods: manually sorted, dual classification with automatic correction, single NN classifier of original feature with manual correction, single NN classifier without correction, and SVM classifier from co-occurrence feature (Fig. 3). The manual correction method (Davis et al. 2004) requires manual removal of false positive images from the automatically classified results for each taxon. The result was further corrected by the probability of detection, which was estimated from Table 2 (last row). The automatic correction method estimates probability of detection and specificity for each taxon from the confusion matrix (last row and last column in Table 1), and used the correction factor discussed in the 'Materials and methods' section.

Except for the copepod category, the manually sorted, manually corrected, and dual-classification curves lie almost on top of each other (Fig. 3). The high agreement between manually corrected and manually sorted results for copepods is due to the incorrect assumption that the human-sorted results were perfect and invariant. For this case, false positive samples were determined using a reference table from manually sorted images (i.e. no variations between manually corrected and manually sorted results) rather than by manually correcting the classification results as discussed by Davis et al. (2004). The high agreement between manually sorted and manually corrected results of copepod abundance is an artifact of this treatment. In fact, among the manually sorted images, there is some overlap between the copepod and 'other' categories due to ambiguity in appearance of some of the 'other' images, which may actually have been copepods oriented in such a way as to make identification by a human difficult.

According to the study by Culverhouse et al. (2003), trained personnel can be expected to achieve 67 to 83% self-consistency in difficult labeling tasks. Our copepod category should belong to this case. That is to say, if a human labels all 20 000+ images a second time, the copepod abundance estimation between the 2 human results is likely to differ. The mean abundance estimation for copepods between automatically classified and manually sorted results is very close. The uncertainty in the manually sorted abundance estimation is comparable to the abundance difference between automatic and manually sorted results.

For taxa in high relative-abundance regions, the 3 automatic classification systems agree very well with each other. However, for taxa having low relative abundance or taxa having low relative-abundance regions, the reduction of abundance error rate is marked (cf. Fig. 4). *Chaetoceros* chains make up <2.5% of total plankton in this tow. The abundance estimation error of the dual-classification system is uniformly <50% along the tow path, which is smaller than the natural variation

for replicate plankton tows (Wiebe & Holland 1968, Davis et al. 2004). In the regions of extremely low relative abundance (e.g. Fig. 3 Hours 8 and 12, *C. socialis* colonies), the dual-classification system estimates the abundance significantly higher than manually sorted or manually corrected abundance.

The reduction in abundance error rates by the dual-classification system compared to the single NN classifier (Davis et al. 2004), the SVM classifier with co-occurrence matrices (COM-SVM, Hu & Davis 2005), and manual correction (Davis et al. 2004) are given in Fig. 4. For copepods, the manually corrected result outperforms other methods. As discussed above, this difference is not significant, due to low confidence of the manually sorted result. For rod-shaped diatom chains, the performances of dual classification, manually corrected, and COM-SVM are very similar. They all outperform the single NN classifier. Dual classification has a significant reduction in abundance error compared to the single NN and COM-SVM, while it is close to the manually corrected results. The same is true for *Chaetoceros socialis* colonies. The performance disagreement happens in the regions of extremely low relative abundance. As discussed by Benfield et al. (1996), these regions could be the limits of the optical sampling method (i.e. high magnification VPR camera used). The performance of 4 different methods on hydroid medusae and marine snow is very close. The dual-classification method performs slightly better for marine snow, while the manually corrected method is better for hydroid medusae.

The advantage of using the dual-classification system is to reduce the false alarm rate of each taxon to a level so low that the variation of specificity for each taxon is low in the whole study region. This makes fully automatic correction possible. The dual-classification system substantially decreases the probability of false alarm, while only slightly reducing the probability of detection. By rejecting a small portion of

Table 2. Confusion matrix of the single LVQ-NN classifier, using leave-one-out method. Images used were the same as those in Table 1. Abbreviations as in Table 1. All data are counts, except in the last row and last column, which are percent values

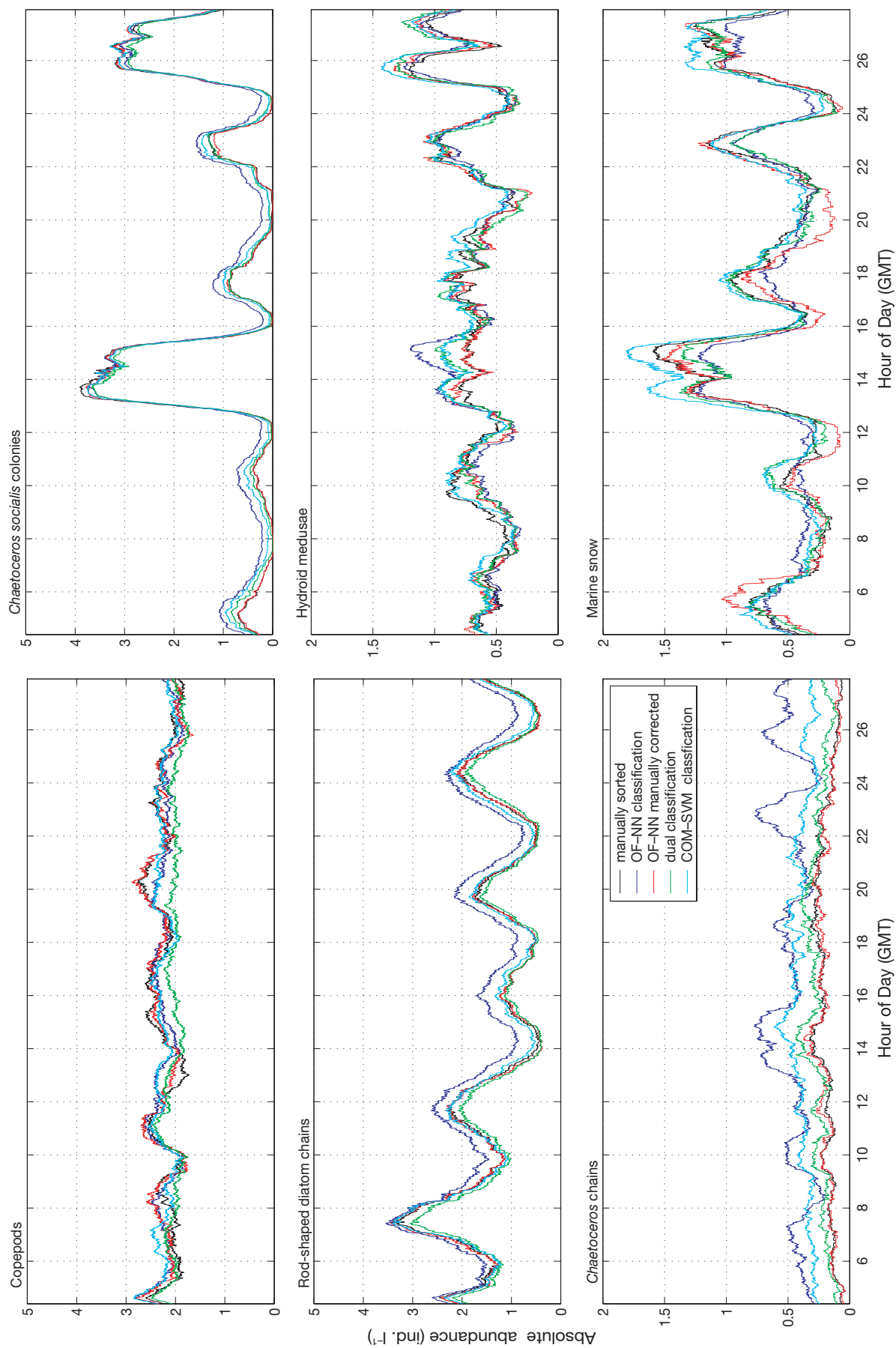|	C1	C2	C3	C4	C5	C6	C7	SP
C1	132	8	16	2	16	7	54	56
C2	6	171	1	4	7	5	20	80
C3	7	1	123	10	23	22	11	62
C4	6	2	14	153	5	20	14	72
C5	5	5	13	4	105	21	9	65
C6	9	3	20	17	34	109	15	53
C7	35	10	13	10	10	16	77	45
P_D	66	86	62	77	53	55	39	NA

Fig. 3. Comparison of 2 automatic classification systems with results classified by human experts. Time-series abundance plots along the tow path are shown for 6 dominant taxa. Data were first binned in 10 s time intervals, and a 1 h smoothing window was applied to the binned data. OF-NN: neural network, based on original features (Davis et al. 2004); COM-SVM: support vector machine, based on co-occurrence matrices (Hu & Davis 2005)
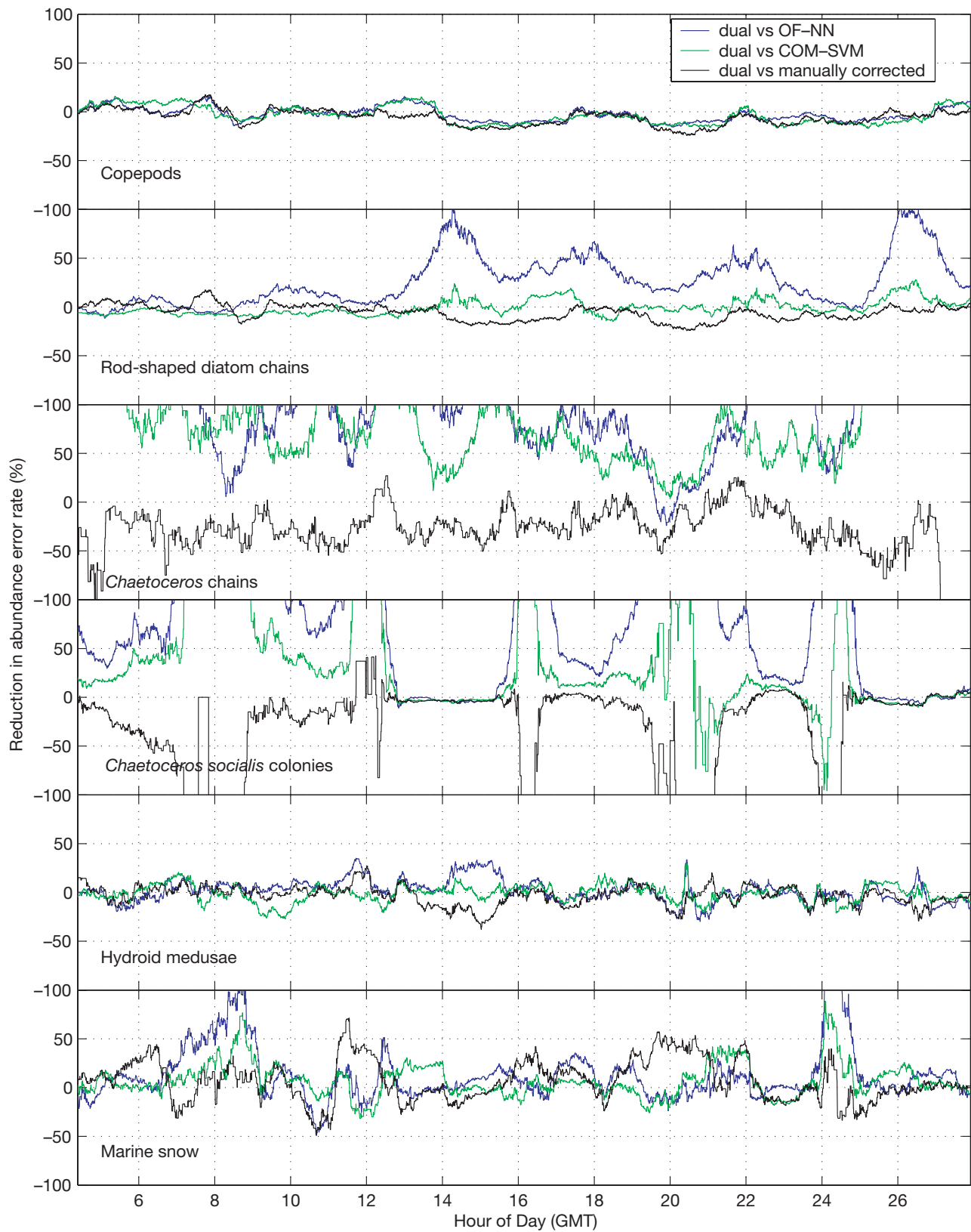
Fig. 4. Reduction in error rate between dual- and single-classification systems for relative abundance estimation. Positive values indicate that the dual-classification system outperforms other methods; negative values indicate the opposite. Dual: dual-classification system; other abbreviations as in Fig. 3

the images as 'unknown', identifications are made by the dual-classification system with higher confidence. Thus, it is not necessary to classify all the images into taxonomic groups to achieve better estimation.

Other approaches to incorporate shape-based features and texture-based features are certainly possible. In the simplest example, all available features are stacked into 1 feature vector and used in training an SVM or LVQ-NN classifier. We have found that such an approach is not efficient, and that the result is almost identical to the COM-SVM method. We have also tried more sophisticated approaches to reduce feature dimension without losing discriminant power, but have thus far met with little success. This approach requires further research.

A dual-classification system utilizes a greater range of variation in feature sets and classifiers. The second classifier provides additional information that the first classifier alone does not possess. It is certainly possible to use 1 type of classifier (e.g. SVM or LVQ-NN) with all types of features or 1 type of feature for both classifiers. However, the variability gained by the dual method using different features and 2 classifiers would be reduced.

The difficulty of general object classification may be overlooked because humans are so good at visual classification of objects. We take for granted our ability to identify facial images without considering the millions of years of evolution involved. On the other hand, a computer is taught in <1 h to identify plankton images that suffer from projection variance, occlusion, non-uniform illumination, and noise, using 200 training images per taxon. The study conducted by Davis et al. (2004) revealed the difficulty level of this data set. They showed that the 90%+ accuracy on a selected subset of these data (Tang et al. 1998) only yielded 60% accuracy on the entire data set. Although humans are able to identify some of the images in this data set with a higher level of taxonomic resolution, in terms of abundance estimation of major taxonomic groups, the dual-classification method presented here yields results almost as accurate as those of human experts.

CONCLUSION

In this paper, we used a dual-classification system by building an SVM classifier on texture-based features (co-occurrence matrices) and an NN classifier on shape-based features (moment invariants, Fourier descriptors, and granulometry) to jointly identify over 20 000 VPR images. A confusion matrix was built from training samples. Sensitivity and specificity of the classification system were calculated from the confusion matrix to correct the abundance estimation. After correction, the dual-classification system can reliably estimate the abun-

dance of a taxon even when its relative abundance is as low as 2.5%. In regions of relatively low abundance, the dual-classification system can reduce the abundance estimation error by 50 to 100% compared with previous methods. Because it is fully automatic, this method can be used for real-time applications, as in our previous methods (Davis et al. 2004, Hu & Davis 2005).

LITERATURE CITED

Ashjian CJ, Davis CS, Gallager 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, Davis CS, Gallager SM, Alatalo P (2005) Characterization of the zooplankton community and size composition in relation to hydrography and circulation in the Sea of Japan. Deep-Sea Res II 52:1363–1392

Ashjian CJ, Gallager SM, Plourde S (in press) Transport of plankton and particles between the Chukchi and Beaufort Seas during summer 2002, described using a Video Plankton Recorder. Deep-Sea Res II

Baldi P, Brunak S (2001) Bioinformatics — the machine learning approach. MIT Press, Cambridge, MA

Benfield MC, Davis CS, Wiebe PH, Gallager SM, Lough RG, Copley NJ (1996) Video Plankton Recorder estimates of copepod, pteropod and larvacean distributions from a stratified region of Georges Bank with comparative measurements from a MOCNESS sampler. Deep Sea Res II 43: 1925–1945

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. Bin Comput Microbiol 5:17–22

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

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, Willaims 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, Gallager SM, Berman MS, Haury LR, Strickler JR (1992a) The Video Plankton Recorder (VPR): design and initial results. Arch Hydrobiol Beih 36:67–81

Davis CS, Gallager SM, Solow AR (1992b) Microaggregations of oceanic plankton observed by towed video microscopy. Science 257:230–232

Davis CS, Gallager 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, Gallager 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, Gallager SM, Hu Q (2005) A three-axis fast-tow digital Video Plankton Recorder for rapid surveys of plankton taxa and hydrography. Limnol Oceanogr Methods 3:59–74

Decoste D, Schölkopf B (2002) Training invariant support vector machines. Mach Learn 46:161–190

Fasham MJR (1978) The statistical and mathematical analysis of plankton patchiness. Oceanogr Mar Biol Annu Rev 16: 43–79

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, Guibert P, Valenta E (1989) The autonomous image analyzer — enumeration, measurement and identification of marine phytoplankton. Mar Ecol Prog Ser 58:133–142

Granlund GH (1972) Fourier preprocessing for hand print character recognition. IEEE Trans Comput 21:195–201

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 Trans 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

Herman AW (1992) Design and calibration of a new optical plankton counter capable of sizing small zooplankton. Deep-Sea Res 39:395–415

Herman AW, Beanlands B, Phillips EF (2004) The next generation of optical plankton counter: the Laser-OPC. J Plankton Res 26:1135–1145

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

Holliday DV, Donaghay PL, Greenlaw CF, McGehee DE, McManus MA, Sullivan JM, Miksis JL (2003) Advances in defining fine- and micro-scale pattern in marine plankton. Aquat Living Resour 16:131–136

Hu MK (1962) Visual pattern recognition by moment invariants. IRE Trans Information Theory 8:179–187

Hu Q, Davis CS (2005) Automatic plankton image recognition with co-occurrence matrics and support vector machine. Mar Ecol Prog Ser 295:21–31

Jeffries HP, Sherman K, Maurer R, Katsinis C (1980) Computer processing of zooplankton samples. In: Kennedy V (ed) Estuarine perspectives. Academic Press, New York, p 303–316

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

Kim H, Howland P, Park H (2005) Dimension reduction in text classification with support vector machines. J Mach Learn Res 6:37–53

Kuhl FP, Giardina CR (1982) Elliptic Fourier features of a closed contour. Comput Vision Graph 18:236–258

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 Trans Syst Man Cyb 34:1753–1762

Mackas DL, Deman KL, Abbott MR (1985) Plankton patchiness: biology in the physical vernacular. Bull Mar Sci 37:652–674

Marine Zooplankton Colloqium (1989) Future marine zooplankton research — a perspective. Mar Ecol Prog Ser 55:197–206

Marine Zooplankton Colloqium (2001) Future marine zooplankton research — a perspective. Mar Ecol Prog Ser 222: 297–308

Matheron G (1975) Random sets and intergral geometry. Wiley, New York

Meijster A, Wilkinson MHK (2002) A comparison of algorithms for connected set openings and closings. IEEE Trans Pattern Anal 24:484–494

Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X (2004) Learning to decode cognitive states from brain images. Mach Learn 57:145–175

Mohan A, Papageorgiou C, Poggio T (2001) Example-based object detection in images by components. IEEE Trans Pattern Anal 23:349–361

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

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

Serra J (1982) Image analysis and mathematical morphology. Academic Press, New York

Simpson R, Culverhouse PF, Ellis R, Willaims R (1991) Classification of *Euceratium* Gran. in neural networks. In: IEEE International Conference on Neural Networks in Ocean Engineering. Institute of Electrical and Electronics Engineers, 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 Trans Syst Man Cyb 28:532–543

Tang X, Stewart WK, Vincent L, Huang H, Marra M, Gallager SM, Davis CS (1998) Automatic plankton image recognition. Artificial Intelligence Rev 12:177–199

Toth L, Culverhouse PF (1999) Three-dimensional object categorisation from static 2D views using multiple coarse channels. Image and Vision Computing 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 algorithms. IEEE Trans Image Process 2:176–201

Wiebe PH, Holland WR (1968) Plankton patchiness: effect of repeated net tows. Limnol Oceanogr 13:315–321

Wiebe PH, Stanton TK, Greene CH, Benfield MC, Sosik HM, Austin T, Warren JD, Hammar T (2002) BIOMAPER II: an integrated instrument platform for coupled biological and physical measurements in coastal and oceanic regimes. IEEE J Ocean Eng 27:700–716

Zahn CT, Roskies RZ (1972) Fourier descriptors for plane closed curves. IEEE Trans Comput 21:269–281