Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Binary SIPPER plankton image classification using random subspace

Feng Zhao *, Feng Lin, Hock Soon Seah

School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

Available online 12 March 2010

Keywords:
Plankton classification
Principal component analysis
Random subspace

ABSTRACT

Plankton recognition plays an important role in the ocean environmental research. In this paper, we propose a random subspace based algorithm to classify the plankton images detected in real time by the Shadowed Image Particle Profiling and Evaluation Recorder. The difficulty of such classification is multifold because the data sets are not only much noisier but the plankton are deformable, projection-variant, and often in partial occlusion. In addition, the images in our experiments are binary, thus are lack of texture information. Using random sampling, we construct a set of stable classifiers to take full advantage of nearly all the discriminative information in the feature space of plankton images. The combination of multiple stable classifiers is better than a single classifier. We achieve over 93% classification accuracy on a collection of more than 3000 images, making it comparable with what a trained biologist can achieve by using conventional manual techniques.

1. Introduction

Plankton including phytoplankton and zooplankton form the base of the food chain in the ocean and are a fundamental component of marine ecosystem dynamics. It is very important but rather difficult to rapidly map plankton abundance together with taxonomic and size composition for ocean environmental research, which can help us understand how climate change and human activities affect the marine ecosystems.

Earlier researchers investigated the temporal and spatial variability in plankton abundance and composition by manually counting the samples collected using traditional methods (e.g., towed nets, pumps, and Niskin bottles), which is laborious and time consuming. To improve the sampling efficiency, some new instruments such as the video plankton recorder (VPR) [1], the HOLOMAR underwater holographic camera system [2], and the shadowed image particle profiling and evaluation recorder (SIPPER) [3] have been developed to continuously sample magnified plankton images in the ocean. These underwater video systems are capable of collecting a large amount of image data over even short periods of time, necessitating an automated pattern recognition system to classify plankton images.

The experimental data sets in this work come from the SIPPER system recently developed by University of South Florida. The SIPPER images differ from those used for most previous research in four aspects: (i) the underwater images are much noisier, (ii) the plankton objects are deformable and often partially occluded, (iii) the images are projection variant, i.e., the images are video records of three-dimensional objects in arbitrary positions and orientations, and (iv) the images in our experiments are binary thus are devoid of texture information. Fig. 1 shows some typical examples to illustrate the diversity of the SIPPER images.

To deal with these difficulties, we combine the general features [4] (e.g., moment invariants [5,6], Fourier descriptors [7,8], and granulometric features [9,10]) with some specific features [11] (e.g., circular projections, boundary smoothness, and object density) to form a more complete description of the binary plankton patterns. The combined feature vector has a high dimensionality, containing much redundant and correlative information. To remove redundancy and reduce noise, we apply the principal component analysis (PCA) [12] technique to compact the combined feature vector, with the eigenvectors corresponding to small eigenvalues removed in the PCA subspace [13]. Since these eigenvectors may encode some useful information for recognition, their removal may introduce a loss of discriminative information, thus affects the overall classification accuracy.

To solve this problem, we propose a novel approach based on random subspace [14]. The random subspace technique has been shown to be very effective for face recognition [15], image retrieval [16], and hyperspectral data classification [17]. In the traditional random subspace method, a number of low-dimensional subspaces are generated by randomly sampling from the original high-dimensional feature space. Finally, multiple classifiers constructed in the random subspaces are combined to make a powerful decision [18,19]. In this work, we do random sampling in the reduced PCA subspace instead and the random subspace is not completely random by fixing the first several dimensions of each subspace as those largest eigenvectors. Moreover, we normalize all types of features with different length and scales in the combined feature vector to the same scale. In doing so, the

* Corresponding author.
E-mail address: zhaofeng@ntu.edu.sg (F. Zhao).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.12.033

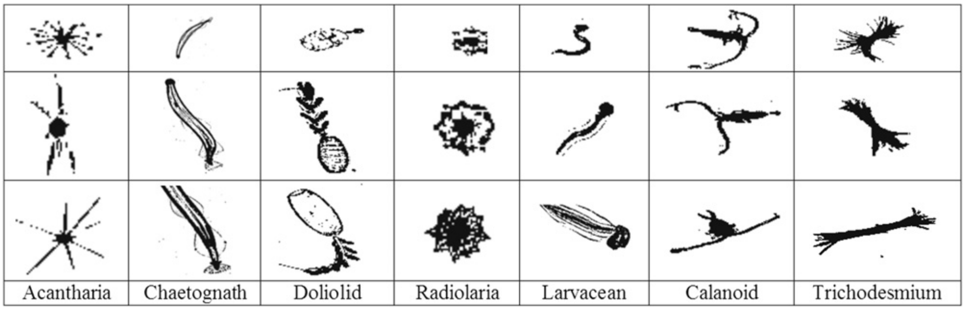

Fig. 1. Examples of seven plankton classes. (All images are rescaled for display purpose.)

constructed classifiers are stable with satisfactory accuracy rates and multiple classifiers cover nearly the entire feature space without losing much discriminative information. Thus, good performance can be achieved. The experiments on seven classes of more than 3000 binary plankton images clearly demonstrate the efficiency and superiority of our algorithm.

The rest of the paper is organized as follows. Section 2 briefly introduces the feature extraction and assessment. Section 3 presents the proposed random subspace based scheme and Section 4 gives its theoretical analysis. Experimental results and discussions are reported in Section 5. Section 6 concludes this paper.

2. Feature extraction and assessment

In selecting and designing features for the binary SIPPER plankton images, we prefer the features that are invariant to irrelative transformations (translation, scale, and rotation), insensitive to noise, and useful for discriminating patterns in different categories.

According to the prior knowledge of the plankton recognition, we choose three most valuable general features, i.e., moment invariants [5,6], Fourier descriptors [7,8] (FD and filled FD), and granulometries [9,10]. Especially, we add three types of structure elements (square, disk, and rhombus) of increasing size to compute the granulometric features since granulometries are relatively robust to noise, occlusion, and projection directions. In order to form a more complete representation of the SIPPER plankton images, we also design a set of specific features [11] such as circular projections (CMS, P_1, P_2, filled P_1, and filled P_2), boundary smoothness, object density, moment ratio, opening distance, relative centroid, and some other geometric features. A brief description of them is given in Table 1, where “Filled” means the features are extracted from the plankton images after applying the flood-filling operation that fills all the internal holes. In the following, we will demonstrate some specific features as examples to show their capability of distinguishing different plankton categories.

2.1. Boundary smoothness

Generally, some plankton objects such as Larvacean have a smooth kernel portion with a long tail. It is difficult to use the Fourier descriptor features to measure the smoothness of the whole object since the smooth portion and the tail portion will be averaged together. To solve this problem, we propose a smoothness measure that only works on the kernel section of the object, i.e., boundary smoothness (M_λ). It is calculated by

$$M_\lambda = \frac{P^2/A}{\pi(\lambda+1)^2/\lambda}, \tag{1}$$

Table 1
Feature description.

Features	Feature length
Boundary smoothness	1
Moment ratio	1
Object density	1
Opening distance	1
Relative centroid	2
Geometric features	6
Moment invariants	7
Granulometries	21
P_1	50
P_2	50
Filled P_1	50
Filled P_2	50
CMS	180
FD	180
Filled FD	180

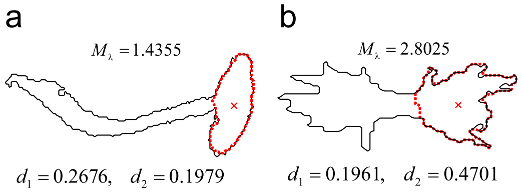
a
b
$M_\lambda = 1.4355$
$M_\lambda = 2.8025$
$d_1 = 0.2676$, $d_2 = 0.1979$
$d_1 = 0.1961$, $d_2 = 0.4701$

Fig. 2. The extracted kernel boundary (marked by the red dashed curves) overlaying on the object boundary with their respective boundary smoothness and relative centroid values: (a) A sample of class Larvacean and (b) a sample of class Trichodesmium. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where P and A are the perimeter and area of the kernel section, $\lambda = a/b$ is the ratio between the length a and width b of the bonding box which encloses the object’s kernel section. If the boundary is a smooth closed curve, we have $M_\lambda \to 1$. Particularly, we have $M_\lambda = 1$ if the boundary is an ellipse curve.

Besides, we extract the relative location of the kernel centroid to the object centroid as a feature (termed as relative centroid). It is defined as

$$d_1 = \sqrt{(x_1 - x_c)^2 + (y_1 - y_c)^2}/w, \tag{2}$$

$$d_2 = \min(x_1, w - x_1 + 1)/h, \tag{3}$$

where (x_1, y_1) are the coordinates of the kernel centroid, (x_c, y_c) are the coordinates of the object centroid, and $[h, w]$ are the height and width of the object.

Fig. 2 shows two object boundary examples of different plankton classes with their respective kernel boundaries, from which we can see that the values of their kernel boundary smoothness and relative centroid are significantly different.

2.2. Object density

To deal with the diversity of the plankton shapes, we define a feature of object density to measure the width of the object's dominant section.

Fig. 3 shows an example of the dominant object pixel-width ($\overline{w_p}$) computation for a plankton sample of class Trichodesmium. In the object pixel-width histogram $L(w)$, the horizontal axis represents the pixel width w and the vertical axis denotes the number of all the segments along the body's principal axis that have the same width. The area distribution of the object pixel-width $S(w)$ represents the area of these segments ($S(w)=L(w)\times w$). The w corresponding to the maximum value in $S(w)$ is considered as the dominant pixel-width of the object, which reflects the largest part of the object. In a similar way, we can compute the object body-width ($\overline{w_b}$) that describes the dominant boundary width of the object, as depicted in Fig. 4.

The object density (D) is defined as the ratio between $\overline{w_p}$ and $\overline{w_b}$ ($D=\overline{w_p}/\overline{w_b}$), which reflects the density property of the binary objects. It is a useful measure for the plankton patterns. For example, the density of class Doliolid is sparse, while the density of class Trichodesmium is dense. As illustrated in Fig. 5, their object density values are very different, which can be used for discriminating the two plankton categories.

All the extracted features have the desirable translation, scale, and rotation invariant property. Every type of feature characterizes the plankton patterns from different aspect. However, since the features are all based on the same plankton object, they may contain much redundant or correlative information. In addition, they have different length (long or short) and scales

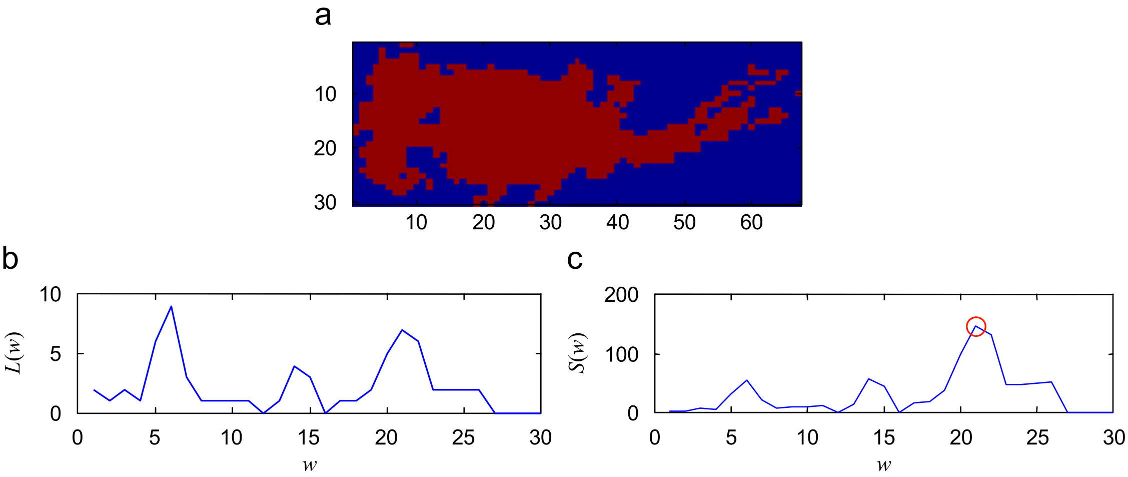

Fig. 3. Computation of the dominant object pixel-width: (a) A rotated binary plankton sample of class Trichodesmium with the horizontal axis corresponding to its principal axis; (b) the histogram of the object pixel-width and (c) the area distribution of the object pixel-width. The dominant object pixel-width is marked by '○' at w=21.

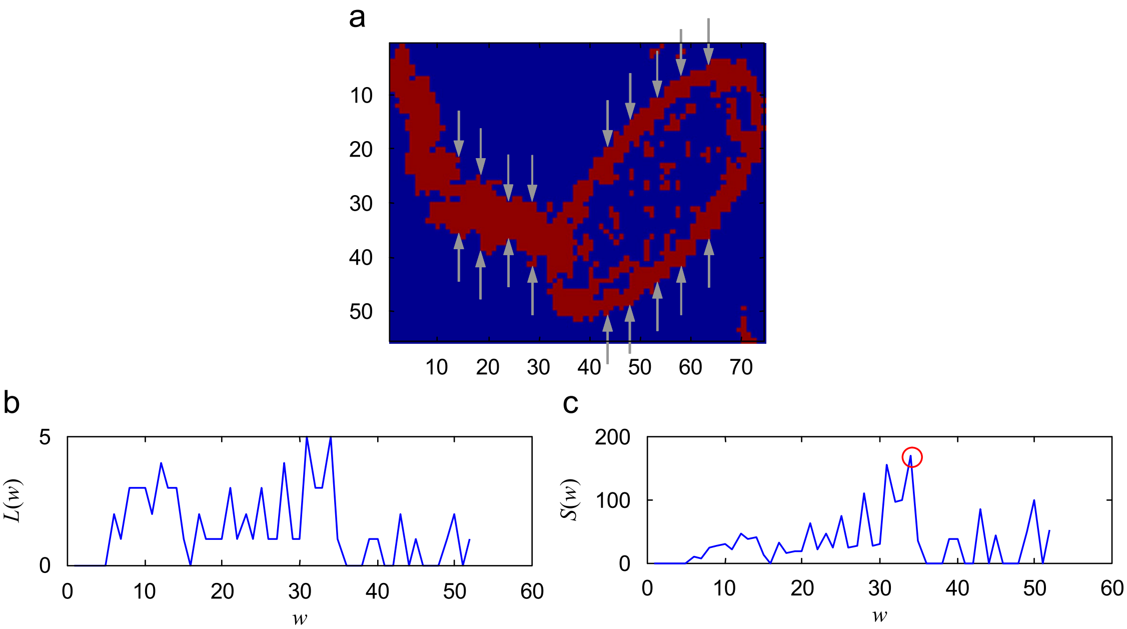

Fig. 4. Computation of the dominant object body-width: (a) A rotated binary plankton sample of class Doliolid, where the gray scale arrows denote the computation of the object body-width values; (b) the histogram of the object body-width and (c) the area distribution of the object body-width. The dominant object body-width is marked by '○' at w=34.

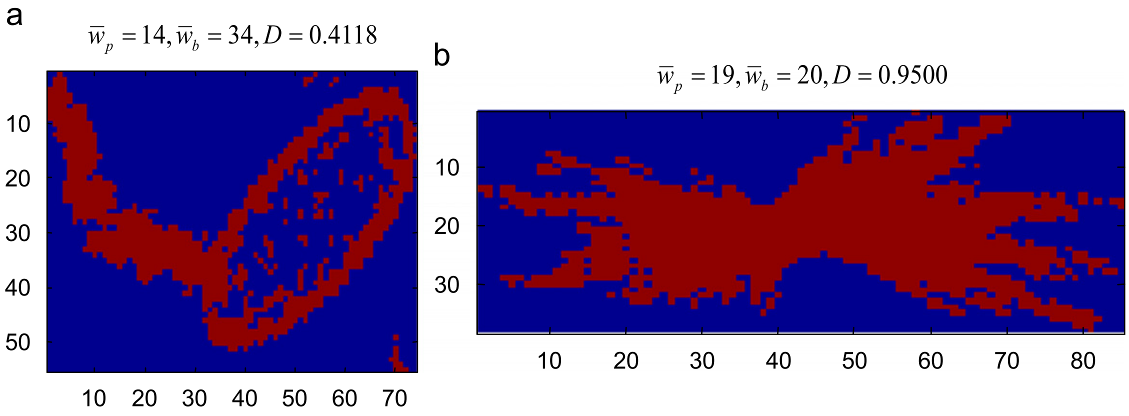

Fig. 5. Examples of two plankton classes with their respective object density values: (a) A rotated binary plankton sample of class Doliolid with a sparse object density and (b) a rotated binary plankton sample of class Trichodesmium with a dense object density.

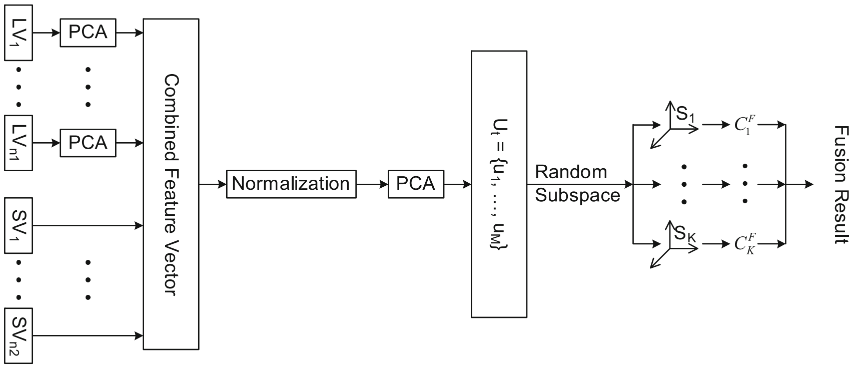

Fig. 6. Multi-feature multi-classifier recognition system using random sampling.

(large or small). To effectively combine these features and take full advantage of them, we propose a two-stage PCA based scheme for feature selection, combination, and normalization (see Fig. 6).

3. Random sampling on combined feature vector

As shown in Table 1, the combined feature vector has a high dimensionality. Normally the high-dimensional feature vector is projected to a low dimensional feature space using PCA to avoid classifier overfitting problem due to the relatively small training set. In order to construct a stable classifier, the eigenvectors with small eigenvalues are usually removed in the PCA subspace. However, eigenvalue is not an indicator of the feature discriminability. Their removal may introduce a loss of useful discriminative information. To solve this problem, we randomly sample a small subset of features to reduce the discrepancy between the training set size and the feature vector length. Using such a random sampling method, we can construct multiple stable classifiers. We then combine these classifiers to construct a more powerful classifier that covers most of the feature space, so less discriminative information is lost.

As illustrated in Fig. 6, our random subspace based approach is designed as follows.

At the training stage,

1. Apply the first-stage PCA to all the long feature vectors (LV_i, $i=1, \ldots, n_1$) such as P_1, P_2, filled P_1, filled P_2, CMS, FD, and filled FD to compute their eigenvectors U_i and eigenvalues λ_i, respectively. The eigenvectors with larger eigenvalues that preserve 98% of the total energy are preserved. The eigenvectors with very small eigenvalues mostly containing noise are removed.
2. For each training sample, project every long feature vector LV_i to the respective eigenvectors (PCA subspace) using $W_i = U_i^T (LV_i - m_i)$, where m_i is the mean of all the i th long feature vectors.
3. Combine the projected feature vectors (W_i, $i=1,\ldots,n_1$) with the short feature vectors (SV_j, $j=1,\ldots,n_2$) such as boundary smoothness, object density, moment invariants, and granulometries to form a new combined feature vector.
4. Normalize every component of the combined feature vector to the same scale.
5. Apply the second-stage PCA to the normalized feature vectors of all the training samples. The first M largest eigenvectors $U_t = \{u_1, \ldots, u_M\}$ that preserve 99% of the total energy are selected as candidates to construct the random subspaces.
6. Generate K random subspaces $\{S_n\}_{n=1}^K$. Each random subspace S_i is spanned by $N_0 + N_1$ dimensions. The first N_0 dimensions are fixed as the N_0 largest eigenvectors in U_t and the other N_1 dimensions are randomly selected from the remaining $M-N_0$ eigenvectors in U_t.
7. Construct K classifiers $\{C_n^F\}_{n=1}^K$ from the corresponding K random subspaces.

At the recognition stage,

1. For each testing image, project its long feature vectors to their respective PCA subspaces.
2. Combine the projected feature vectors with the short feature vectors to form a new combined feature vector.

3. Normalize every component of the combined feature vector to the same scale.
4. Project the normalized feature vector to each of the K random subspaces and feed them to the K corresponding classifiers in parallel.
5. Combine the outputs of the K classifiers using a fusion rule to make the final decision.

Compared with the traditional random subspace method that samples the original feature vector directly, our algorithm samples in the PCA subspace. The high dimension of the feature space is first greatly reduced without losing discriminative information. After doing PCA projection, the features on different eigenvectors are more independent. Better accuracy can be achieved if different random subspaces are more independent from one another.

Secondly, the random subspace is not completely random. The random subspace dimension is fixed as N_0+N_1, which is determined by the training set to make a single classifier stable. In each random subspace, the first N_0 dimensions are fixed as the N_0 largest eigenvectors and the other N_1 dimensions are randomly selected from the remaining $M-N_0$ eigenvectors. The individual classifier constructed in each random subspace has a satisfactory accuracy since the first N_0 dimensions encode much information. If they are not included, the accuracy of each individual classifier may be low and the fusion method to combine these weak classifiers will be more complicated. In addition, the N_1 random dimensions cover most of the remaining small eigenvectors. Thus, our approach makes use of nearly all the discriminative information in the feature space. The ensemble classifiers also have a certain degree of error diversity. Good performance can be achieved using simple fusion rules such as majority voting [20].

Thirdly, we use the first-stage PCA to compact every long feature vector by removing the redundant information. The noise within every long feature vector is removed as well. We then use the second-stage PCA to compact the combined feature vector by removing the correlative information among different types of features since they are all based on the same object shape. Furthermore, because the original feature vectors have different length and scales, the scale of the projected long feature vectors and the short feature vectors can be very different. One may overwhelm the other. We normalize all types of features in the combined feature vector to the same scale according to their respective mean value and standard deviation, thus make use of the contributions of all the features (large or small). Significant improvement can be achieved after normalization.

4. Theoretical analysis on random subspace

4.1. Theoretical study

Let us assume that each sample $(y,\boldsymbol{x})$ in the training set L' is independently drawn from the underlying probability distribution P, where y is the voting indicator of sample $\boldsymbol{x}$. For example, $y=(1\ 0\ \ldots\ 0)$ indicates that $\boldsymbol{x}$ belongs to class 1, $y=(0\ 1\ \ldots\ 0)$ indicates that $\boldsymbol{x}$ belongs to class 2, and so on. The corresponding feature space is denoted by F'. $\varphi(\boldsymbol{x},F)$ is an individual predictor (classifier) constructed in the feature subspace F, where F is generated by randomly sampling on the full feature space F' using random subspace method. Then the aggregated classifier (combination of multiple predictors) is

$$\varphi_A(\boldsymbol{x},P)=E_F\varphi(\boldsymbol{x},F), \tag{4}$$

where $E_F\varphi(\boldsymbol{x},F)$ is the expectation of $\varphi(\boldsymbol{x},F)$ over F.

Let $(Y,\boldsymbol{X})$ be the random variables denoting a testing sample drawn from the identical probability distribution P, and independent of the training set L', where Y is the voting indicator of sample $\boldsymbol{X}$. The average classification error by the individual classifier $\varphi(\boldsymbol{x},F)$ is

$$e=E_FE_{Y,\boldsymbol{X}}(Y-\varphi(\boldsymbol{X},F))^2. \tag{5}$$

The corresponding classification error by the aggregated classifier $\varphi_A(\boldsymbol{x},P)$ is

$$e_A=E_{Y,\boldsymbol{X}}(Y-\varphi_A(\boldsymbol{X},P))^2. \tag{6}$$

Applying the inequality,

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}z_i^2}\geq\frac{1}{N}\sum_{i=1}^{N}z_i,\ \Rightarrow\ \frac{1}{N}\sum_{i=1}^{N}z_i^2\geq\left(\frac{1}{N}\sum_{i=1}^{N}z_i\right)^2, \tag{7}$$

we have

$$E_F\varphi^2(\boldsymbol{X},F)\geq(E_F\varphi(\boldsymbol{X},F))^2, \tag{8}$$

$$E_{Y,\boldsymbol{X}}E_F\varphi^2(\boldsymbol{X},F)\geq E_{Y,\boldsymbol{X}}(E_F\varphi(\boldsymbol{X},F))^2=E_{Y,\boldsymbol{X}}\varphi_A^2(\boldsymbol{X},P). \tag{9}$$

Thus,

$$\begin{aligned}e&=E_FE_{Y,\boldsymbol{X}}(Y-\varphi(\boldsymbol{X},F))^2\\&=E_{Y,\boldsymbol{X}}E_FY^2-2E_{Y,\boldsymbol{X}}YE_F\varphi(\boldsymbol{X},F)+E_{Y,\boldsymbol{X}}E_F\varphi^2(\boldsymbol{X},F)\\&=E_{Y,\boldsymbol{X}}Y^2-2E_{Y,\boldsymbol{X}}Y\varphi_A(\boldsymbol{X},P)+E_{Y,\boldsymbol{X}}E_F\varphi^2(\boldsymbol{X},F)\\&\geq E_{Y,\boldsymbol{X}}Y^2-2E_{Y,\boldsymbol{X}}Y\varphi_A(\boldsymbol{X},P)+E_{Y,\boldsymbol{X}}\varphi_A^2(\boldsymbol{X},P)\\&=E_{Y,\boldsymbol{X}}(Y-\varphi_A(\boldsymbol{X},P))^2\\&=e_A.\end{aligned} \tag{10}$$

From the inequality of $e_A\leq e$, we can see that the mean-squared error of the aggregated classifier is smaller than the average mean-squared error of the individual classifier. That is, the ensemble of multiple classifiers reduces the classification error. How much improvement we can get depends on the difference between the two terms: $(E_F\varphi(\boldsymbol{x},F))^2$ and $E_F\varphi^2(\boldsymbol{x},F)$. The more diverse $\varphi(\boldsymbol{x},F)$ is, the better performance the aggregated classifier will achieve. The constructed individual classifiers have large error diversity, because the random subspaces extract discriminative information from different portions of the feature space. Therefore, the ensemble of these classifiers will lead to an improved system performance.

Here we assume that the average performance of all the individual classifiers $\varphi(\boldsymbol{x},F)$ trained in each feature subspace is similar to a single classifier trained in the full feature space. This can be true when the size of the feature subspace is sufficiently large. Even when this is not true, the drop of accuracy for each individual classifier may be well compensated in the aggregation process.

4.2. The number of random subspaces

Each random subspace is spanned by N_0 + N_1 dimensions. The first N_0 dimensions are fixed and the other N_1 dimensions are randomly selected. The question is how to determine the number of random subspaces (classifiers) to obtain an acceptable performance for the plankton classification, i.e., how many times do we need for the random selections from the remaining eigenvectors? In the following, we will propose a probability model to compute the value of K (the number of classifiers) at the training stage.

The criterion is to make the best of the remaining $M-N_0$ eigenvectors in U_t, i.e., to cover the remaining feature space as much as possible.

Let us assume that we have s ($s=M-N_0$ in this work) eigenvectors in the remaining feature space and we randomly select t ($t=N_1$ in this work) eigenvectors from the s eigenvectors each time. After selection for r times, the probability of occurrence for s_1 eigenvectors selected from the s eigenvectors is computed by

$$p(s_1) = p(s_1 \text{ selected}) = p(s - s_1 \text{ not selected}) = \binom{s-s_1}{s} \cdot (1-t/s)^{(s-s_1)r}, \qquad (11)$$

where $(1-t/s)$ denotes the probability of some eigenvector not selected for some time, $(1-t/s)^r$ denotes the probability of some eigenvector not selected for r times, and $(1-t/s)^{(s-s_1)r}$ denotes the probability of some $(s-s_1)$ eigenvectors not selected for r times.

The usage rate of the eigenvectors after the r random selections is modeled by the probability function $f(r)$,

$$f(r) = \sum_{s_1=t}^{s} \frac{p(s_1)\cdot s_1}{s} = \sum_{s_1=t}^{s} \binom{s-s_1}{s} \cdot \left(1-\frac{t}{s}\right)^{(s-s_1)r} \cdot \frac{s_1}{s}. \qquad (12)$$

The larger the r is, the higher the usage rate $f(r)$ will be.

Given a certain usage rate, the corresponding r is considered to be the number of random subspaces (classifiers), i.e., K. In this work, we generate K=20 random subspaces for experiments, which corresponds to use 80% of the remaining feature space. Note that this does not mean that 20 is necessary or sufficient, but simply that it seems reasonable. A larger K may improve the accuracy, but increase the system burden.

5. Experiments

Experiments are conducted on seven classes of 3119 binary SIPPER plankton images from the Gulf of Mexico. The data set consists of 131 Acantharia, 172 Chaetognath, 450 Doliolid, 485 Radiolaria, 529 Larvacean, 563 Calanoid, and 789 Trichodesmium. All the images were manually classified by marine scientists. In our experiments, the Gaussian minimum error classifier [21] is adopted to classify the plankton images and the 10-fold cross-validation technique [22] is employed to evaluate the performance of our algorithm.

We first apply the PCA to the seven long feature vectors. Depending on their energy distributions shown in Fig. 7, we preserve the top 3–5 eigenvectors with larger eigenvalues for every feature because they encode most of the discriminating information. Table 2 gives the comparison between the original and preserved length of every long feature vector. Since each type of feature (long or short) characterizes the plankton shape from a different aspect, we combine these feature vectors and normalize them to the same scale to obtain a better performance.

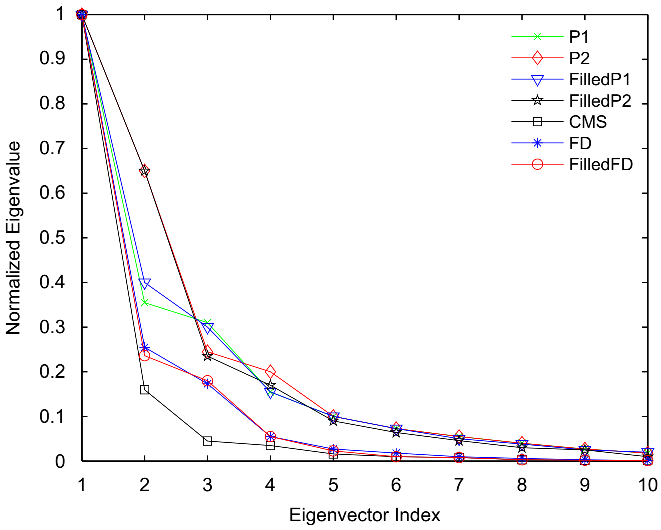

Fig. 7. The energy distributions of top 10 eigenvectors after applying PCA to the long feature vectors.

Table 2
Comparison between the original and preserved feature length.

Features	Original length	Preserved length
P_1	50	5
P_2	50	5
Filled P_1	50	5
Filled P_2	50	4
CMS	180	4
FD	180	3
Filled FD	180	3

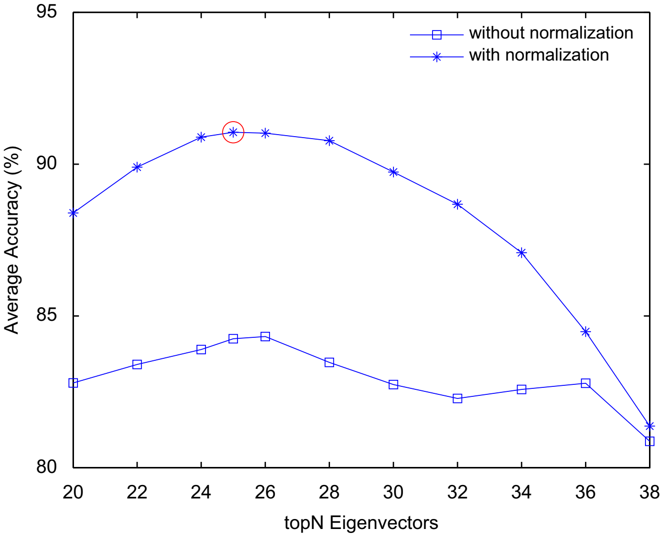

Fig. 8. The average accuracy of a single classifier using different number of largest eigenvectors in the reduced PCA subspace. The red circle mark denotes the best accuracy rate of 91.05% for a single classifier. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 8 shows the average accuracy of a single classifier constructed in the PCA subspace using different number of largest eigenvectors. We can see a remarkable improvement of the classification accuracy since all the features after normalization can contribute to the classification. Also, we observe that a single classifier has the best accuracy of 91.05% using the largest 25 eigenvectors $\{u_1, \ldots, u_{25}\}$, which seems to be suitable to construct a stable classifier for our experimental data sets. In the following experiments, we choose 25 as the dimension of the random subspaces to construct a number of individual classifiers.

First, for each random subspace, we randomly select its 25 dimensions from the preserved eigenvectors in U_t. An individual classifier is then trained on the selected eigenvectors. Fig. 9 demonstrates the result of combining 20 individual classifiers using majority voting. We can see that with conventional random sampling method, the accuracy of every individual classifier is low, ranging from 50% to 80%. However, these weak classifiers are greatly enforced with majority voting, and 88% accuracy is achieved. The result shows that the individual classifiers constructed in different random subspaces are complementary of one another.

A better approach to improve the overall performance of the combined classifier is to increase the accuracy of every individual weak classifier. Toward this, we fix the first 15 dimensions of each random subspace as the 15 largest eigenvectors, and randomly select the other 10 dimensions from the remaining eigenvectors in U_t. As illustrated in Fig. 10, the individual classifiers constructed in the random subspaces are improved significantly. They are comparable to the single classifier using the largest 25 eigenvectors and have similar accuracies since much information is contained in the first 15 largest eigenvectors $\{u_1, \ldots, u_{15}\}$. The results also indicate that the other 10 eigenvectors with large eigenvalues $\{u_{16}, \ldots, u_{25}\}$ are not more discriminative than those smaller eigenvectors. In Table 3, we report the mean accuracy and standard deviation of these individual classifiers. These classifiers are complementary of one another, so better performance can be achieved when they are combined. The result of combining 20 individual classifiers using majority voting is shown in Fig. 10. It demonstrates that the combination of multiple stable classifiers outperforms the optimal single classifier. Furthermore, our random subspace based method has a superior performance than the NMDEE method in [13]. Increasing classifier number and using more complicated fusion rules may further improve the performance at the expense of computation time.

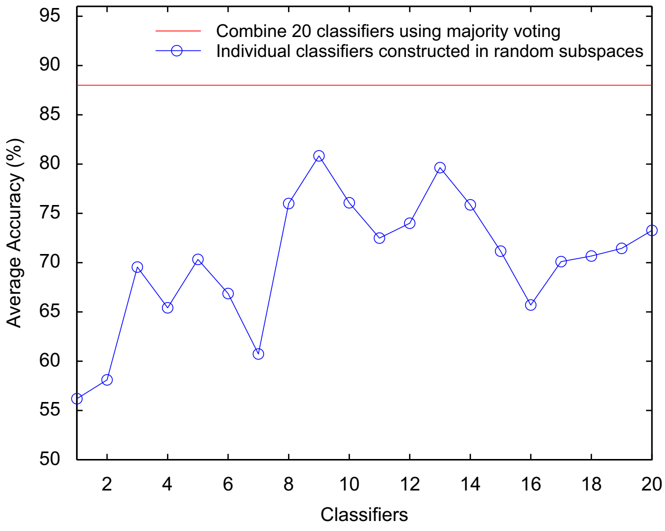

Fig. 9. The average accuracy of combining 20 individual classifiers constructed in the random subspaces using majority voting. All the 25 dimensions of each random subspace are randomly selected.

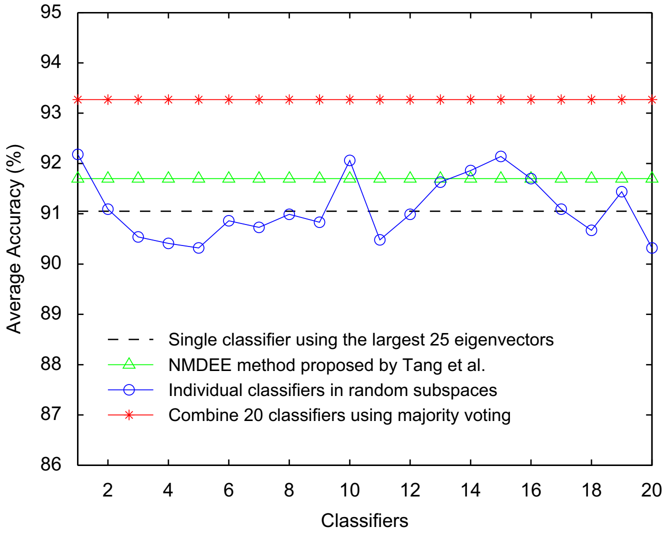

Fig. 10. The average accuracy of combining 20 individual classifiers constructed in the random subspaces using majority voting. The first 15 dimensions of each random subspace are fixed as the 15 largest eigenvectors, and the other 10 dimensions are randomly selected.

The average accuracy rates combining different number of individual classifiers generated by the random subspace method are reported in Table 4. In general, using more classifiers (a larger value of K) leads to a better accuracy rate. When the classifier number is enough to a certain extent, the overall performance may be quite stable. Even if we increase the classifier number, the accuracy will not be improved much, but increase the system burden. The classifier number will be a trade-off between the overall accuracy and the computational cost.

Table 5 shows the confusion matrix of our approach from a 10-fold cross-validation experiment. The overall average accuracy is 93.27%. We have over 90% accuracy on most plankton classes except for Acantharia and Chaetognath. The reason is that we have fewer samples of the two classes.

Table 4
The average accuracy rates combining different number of classifiers.

Classifier number (K)	5	10	20	30	50	100
Average Accuracy (%)	91.66	92.38	93.27	93.95	94.31	94.63
Mean accuracy (%)			93.37			
Standard deviation (%)			1.16			

Table 5
Confusion matrix of random subspace method from a 10-fold cross-validation on 3119 SIPPER plankton images using majority voting.

Class ID	1	2	3	4	5	6	7	Accuracy (%)
1	**115**	0	4	2	0	10	0	87.79
2	0	**147**	15	0	8	0	2	85.47
3	0	5	**416**	19	8	0	2	92.44
4	16	0	16	**448**	0	0	5	92.37
5	1	3	25	0	**491**	1	8	92.82
6	4	0	0	0	3	**544**	12	96.63
7	0	1	9	10	10	11	**748**	94.80
Average accuracy				**93.27%**				

The seven plankton classes are numbered from 1 to 7 in turn.

Table 3
The mean accuracy rate and the standard deviation of the 20 individual classifiers constructed by random subspace method.

Classifier	1	2	3	4	5	6	7	8	9	10
Accuracy (%)	92.18	91.09	90.54	90.41	90.32	90.86	90.73	90.99	90.83	92.06
Classifier	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Accuracy (%)	90.48	90.99	91.63	91.86	92.14	91.70	91.09	90.67	91.44	90.32
Mean accuracy (%)						91.12				
Standard deviation (%)						0.63				

6. Conclusion

In this paper, we have developed a random subspace based approach for the integration of multiple features for binary SIPPER plankton classification. Integration at the feature level conveys the richest information, but it is more difficult because: (i) different types of features are incompatible in length and scale, and (ii) the combined feature vector has a higher dimensionality. Our approach overcomes both problems. Using random sampling, we construct a set of stable classifiers preserving nearly all the discriminative information in the high-dimensional feature space. The combination of multiple classifiers produces a much better performance than a single classifier. Experimental results on a large data set show that our algorithms can effectively classify binary plankton images with a high accuracy rate acceptable for automatic plankton survey system.

Acknowledgments

The authors would like to thank the researchers (S. Samson and A. Remsen) at University of South Florida, USA, who provided the plankton data for this study.

References

[1] C. Davis, S. Gallager, N. Berman, L. Haury, J. Strickler, The Video Plankton Recorder (VPR): design and initial results, Arch. Hydrobiol. Beith. 36 (1992) 67–81.
[2] J. Watson, G. Craig, V. Chalvidan, J.P. Chambard, A. Diard, G.L. Foresti, B. Forre, S. Gentili, P.R. Hobson, R.S. Lampitt, P. Maine, J.T. Malmo, H. Nareid, G. Pieroni, S. Serpico, K. Tipping, A. Trucco, High resolution *in situ* holographic recording and analysis of marine organisms and particles (HOLOMAR), in: Proceedings of the IEEE International Conference on OCEAN'98, 1998, pp. 1599–1604.
[3] S. Samson, T. Hopkins, A. Remsen, L. Langebrake, T. Sutton, J. Patten, A system for high-resolution zooplankton imaging, IEEE J. Ocean. Eng. 26 (4) (2001) 671–676.
[4] X. Tang, W. Stewart, L. Vincent, H. Huang, M. Marra, S. Gallager, C. Davis, Automatic plankton image recognition, Artif. Intell. Rev. 12 (1–3) (1998) 177–199.
[5] M. Hu, Visual pattern recognition by moment invariants, IRE Trans. Inf. Theory IT-8 (1962) 179–187.
[6] T.H. Reiss, The revised fundamental theorem of moment invariants, IEEE Trans. Pattern Anal. Mach. Intell. 13 (8) (1991) 830–834.
[7] H. Kauppinen, T. Seppanen, M. Pietikainen, An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification, IEEE Trans. Pattern Anal. Mach. Intell. 17 (2) (1995) 201–207.
[8] E. Persoon, K.S. Fu, Shape discrimination using Fourier descriptors, IEEE Trans. Syst. Man Cybern. 7 (1977) 170–179.
[9] G. Ayala, J. Domingo, Spatial size distributions: applications to shape and texture analysis, IEEE Trans. Pattern Anal. Mach. Intell. 23 (12) (2001) 1430–1442.
[10] G. Matheron, Random Sets and Integral Geometry, John Wiley and Sons, New York, 1975.
[11] F. Zhao, Binary plankton recognition using random sampling, Ph.D. Dissertation, The Chinese University of Hong Kong, May 2006.
[12] I.T. Jolliffe, Principal Component Analysis, chap.13, second ed., Springer-Verlag, 2002.
[13] X. Tang, F. Lin, S. Samson, A. Remsen, Binary plankton image classification, IEEE J. Ocean. Eng. 31 (3) (2006) 728–735.
[14] T.K. Ho, The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (8) (1998) 832–844.
[15] X. Wang, X. Tang, Random sampling LDA for face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 259–265.
[16] D. Tao, X. Tang, Random sampling based SVM for relevance feedback image retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 647–652.
[17] J. Ham, Y. Chen, M.M. Crawford, J. Ghosh, Investigation of the random forest framework for classification of hyperspectral data, IEEE Trans. Geosci. Remote Sensing 43 (3) (2005) 492–501.
[18] T.K. Ho, J.J. Hull, S.N. Srihari, Decision combination in multiple classifier systems, IEEE Trans. Pattern Anal. Mach. Intell. 16 (1) (1994) 66–75.
[19] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.
[20] L. Lam, S.Y. Suen, Application of majority voting to pattern recognition: an analysis of its behavior and performance, IEEE Trans. Syst. Man Cybern. Part A 27 (5) (1997) 553–568.
[21] R. Duda, P. Hart, D. Stork, Pattern Classification, chap.2, second ed., John Wiley and Sons, Inc., 2001.
[22] A. Webb, Statistical Pattern Recognition, chap.8, second ed., John Wiley and Sons, 2002.

Feng Zhao received the B.Eng. degree from the University of Science and Technology of China, Hefei, in 2000, and the M.Phil. and Ph.D. degrees from the Chinese University of Hong Kong, in 2002 and 2006, respectively. Since October 2007, he has been with the Emerging Research Lab in School of Computer Engineering, Nanyang Technological University, where he is currently a Research Fellow. From June 2006 to September 2007, Dr. Zhao worked as a Postdoctoral Fellow in the Department of Information Engineering, the Chinese University of Hong Kong. His research interests include fingerprint recognition, plankton recognition, biomedical image processing, and pattern recognition.

Feng Lin is an associate professor in School of Computer Engineering, Nanyang Technological University (NTU), Singapore. His research interests include biomedical imaging, graphics and visualization, bioinformatics, and high-performance computing.

Hock Soon Seah is a Professor and Director of the gameLAB in School of Computer Engineering, Nanyang Technological University (NTU), Singapore. He holds a first-class-honor Bachelor degree in Electrical Engineering, a Master degree in Computing Science, and a Ph.D. degree in Computer Graphics.