Environmental Science
Processes & Impacts

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

rsc.li/process-impacts

We present a computer-microscope system for rapid, accurate recognition and classification of microalgae using image processing techniques such as segmentation, shape features extraction, pigment signature determination and neural network grouping. Microalgae, when growing massively, may produce harmful effects on marine or freshwater ecology and fishery resources, hence real-time monitoring of their presence is critical and essential for the proper management of any water body. Our system attained 98,6% accuracy from a set of 53,869 images of 23 different microalgae representing the major algal phyla and could be useful for an appropriate and effective water resource management.

Journal Name

RSCPublishing

ARTICLE

Water monitoring: automated and real time identification and classification of algae using digital microscopy

Cite this: DOI: 10.1039/xoxxoooooox

Primo Coltelli,[a] Laura Barsanti,[b] Valtere Evangelista,[b] Anna Maria Frassanito,[b] and Paolo Gualtieri,*[b]

Received ooth January 2014,
Accepted ooth January 2014

DOI: 10.1039/xoxxoooooox

www.rsc.org/

Microalgae are unicellular photoautotrophs that grow in any habitat from fresh and saline water bodies, to hot springs and ice. Microalgae can be used as indicators to monitor water ecosystem conditions. These organisms react quickly and predictably to a broad range of environmental stressors, thus providing early signals of changing environment. When growing massively, microalgae may produce harmful effects on marine or freshwater ecology and fishery resources. Rapid and accurate recognition and classification of microalgae is one of the most important issues in water resource management. In this paper, a methodology for automatic and real time identification and enumeration of microalgae by means of image analysis is presented. The methodology is based on segmentation, shape features extraction, pigment signature determination and neural network grouping; it attained 98,6% accuracy from a set of 53,869 images of 23 different microalgae representing the major algal phyla. In our opinion this methodology partly overcomes the lack of automated identification systems and is on the forefront of developing a computer-based image processing technique to automatically detect, recognize, identify and enumerate microalgae genera and species from all the divisions. This methodology could be useful for an appropriate and effective water resource management.

Introduction

Environmental monitoring can be defined as the systematic sampling of air, water, soil, and biota in order to observe and study the environment, as well as to derive knowledge from this process [1, 2, 3, 4]. Monitoring of environmental water quality is essential for the proper management of any water body, since a safe, clean water supply is critical for many important ecosystem services. Irrespective of specific uses, the necessity of a sustainable management of inland and coastal water leads to the requirement of control systems to detect different chemical or biological compounds in situ at very low concentrations in order to react quickly for limiting impact on natural surface and ground waters[5].

General water quality parameters including pH, salinity, temperature, flow and turbidity are commonly monitored using on-line instrumentation[6]. Still, to provide a complete spectrum of information for appropriate water management, biological parameters should be assessed, because they reflects the overall water quality, integrating the effects of physical and chemical changes over time. Responses to changes in water quality occur over different time scales for primary producers (i.e. algae and phototrophic bacteria) and consumers, as primary producers often respond to changes earlier. Since algae are the first trophic level, any disturbances could be reported to upper levels. Algae are valuable indicators of ecosystem conditions because they respond strongly and predictably both in species composition and densities to a wide range of water conditions due to changes in water chemistry (nutrient enrichment, organic contamination, changes in pH or conductivity as well as increases in suspended sediments, pesticides and many other contaminants). When contamination occurs, algae are among the first biological organisms to respond. This is a result of their nutrient needs, short lifespan, which averages about 6 to 8 weeks, and generation times ranging from <1 day to a few days. Those species that cannot tolerate the water quality changes will be replaced by species better suited to the new water quality conditions, resulting in an altered taxonomic composition of the algal community[7].

At low numbers, algae cause no problems and are, in fact, a natural part of a water body. Occasionally, however, algae can grow very fast or 'bloom' and accumulate into dense visible patches at the surface of the water. These blooms can become a

[a] Istituto di Scienza e Tecnologia Informazione, CNR, Via Moruzzi 1, 56124 Pisa, Italy.
[b] Istituto di Biofisica, CNR, Via Moruzzi 1, 56124 Pisa, Italy. E-mail: paolo.gualtieri@cnr.it; Fax: +39 050 3152760;Tel:+39 050 3153026.

serious public health and environmental problem in many waterways. As algae die and decompose, high levels of organic matter deplete the water of available oxygen, causing the death of other organisms. Algal blooms can be driven by low water flows and high levels of available nutrients, which promote their growth. Of primary importance from an ecological and public health perspective is the abundance of nutrients containing nitrogen (N) and phosphorus (P) that flow into lakes, reservoirs, streams and rivers resulting in eutrophic conditions[8]. The N:P ratio often determines which algae genera are dominant, present or absent in these nutrient-affected water bodies. When N:P ratios are high, chlorophytes, along with diatoms, are often the dominant genera. Because of their nitrogen-fixing ability, cyanobacteria blooms usually occur when the N:P ratio is low, with phosphorus as the limiting factor for their growth and reproduction. In high densities, cyanobacteria are an undesirable component of freshwater ecosystems because they can produce hepatotoxins and neurotoxins that are ecological and public health concerns. Toxin producing blooms may disrupt lake food webs by killing fish, birds and zooplankton. Toxic blooms can also restrict recreation like swimming, fishing and pet-related activities [9].

For monitoring purposes, phytoplankton analysis of water samples collected from sea, lakes, streams and other water bodies is therefore a valuable assessment tool to determine the diversity and density of algal species and provide potentially useful early warning signs of deteriorating conditions [10]. Algae show a wide variety of size, shape, texture and colors and all these characteristics are routinely used for taxonomical recognition. Conventional identification by means of microscopy is time consuming and researchers must have abundant taxonomy competence and experience of classification to achieve reliable results.

Several systems have been developed to automate the analysis and classification of algae images [11,12,13] and several kinds of automatic analysis and algae identification methods have been set up. They include for example methods based on algal cell morphology identification[14], absorption spectroscopy[15], fluorescence spectroscopy[16, 17, 18], high performance liquid chromatography [19], flow cytometry [20,21], and molecular biology (i.e. gene probe method), [22, 23, 24, 25, 26, 27, 28, 29, 30]. Most efforts are limited to some specific types of algae only, and therefore limited in their applicability.

As alternative to the previously cited techniques, we present a software methodology coupled to a motorized microscope set-up that offers a reliable, real time acquisition, detection, recognition, and enumeration of algal species in multialgal field samples. This software methodology aims to improve accuracy and reliability of the methodology already described in Coltelli *et al*., (2013) [31, 32]. It combines robust image segmentation, shape features extraction, in-focus algae detection and recognition. Though at the present stage of development our methodology cannot be considered ready for field analysis applications, still it is very promising for future automated systems for environmental monitoring and protection of public water supplies. In the following, the term "algae" will be used in place of the term "microalgae".

Experimental

Algae Samples and Cultures

The system was tested on samples obtained by mixing freshwater algae cultivated under controlled conditions in our laboratory, (Table 1). Since samples may contain amount of large non-algal particles and cell clusters, they are filtered through cheesecloth. Cell concentration was assessed by using a Thoma counting chamber. The density of the mixed sample was chosen so that the algae on the slide do not overlap. No fixatives were used. The marine chlorophycean *Tetraselmis suecica* was used to show the in-focus and out-of-focus cells selection.

Table 1 List of algal strains.

Name	Phylum	SOM group
Cyanothece sp.	Cyanobacteria	2
Lyngbya sp.	Cyanobacteria	22
Nostoc commune	Cyanobacteria	17
Closterium sp.	Charophyta	5
Cosmarium laeve	Charophyta	24
Mesostigma viride	Charophyta	11
Haematococcus lacustris	Chlorophyta	8
Pediastrum duplex	Chlorophyta	19
Scenedesmus quadricauda	Chlorophyta	13
Selenastrum gracile	Chlorophyta	3
Tetraselmis suecica	Chlorophyta	
Cryptomonas ovata	Cryptophyta	23
Chroomonas sp.	Cryptophyta	10
Euglena ehrenbergii	Euglenozoa	16
Euglena acus	Euglenozoa	7
Euglena gracilis	Euglenozoa	12
Phacus sp.	Euglenozoa	6
Trachelomonas sp	Euglenozoa	21
Cyanophora paradoxa	Glaucophyta	9
Gymnodinium acidotum	Myzozoa	18
Gomphonema parvulum	Ochrophyta *(girdle view)*	14
Gomphonema parvulum	Ochrophyta *(valve view)*	15
Synura uvella	Ochrophyta	20
Ochromonas danica	Ochrophyta	1
Pennate diatom	Ochrophyta	4

Operating Platform for Digital Microscopy

The hardware platform used to perform *in vivo* and real time image measurements consists of a Zeiss Axioplan microscope (Zeiss, Germany), with a 100 W stabilized tungsten-filament lamp, and a 40x (N.A. 0.75) planapochromatic objective. The microscope is equipped with a manual and motorized mechanical 75x30 mm scanning stage (Märzhäuser, Wetzlar, Germany) with a high-resolution stepper motor controller (minimum step size 0,05 µm) plugged into the computer bus. A digital color CCD camera (Basler scA160028fm/fc, Basler, Germany) equipped with a IEEE 1394b interface was mounted in the TV microscope path. The resolution of the original image is 1628 x 1236 pixels. The personal computer PC is Intel core i7-2600 (3,39 GHz), equipped with an 2 TB HD, 16GB RAM, and Windows 7 operating system.

Image Analysis

Image processing and pattern recognition were performed using MATLAB R2009b software with home-made routines. The graphical user-friendly interface of the system allows image acquisition, image processing, image segmentation, feature extraction and automated algae classification and enumeration. These images were stored into a database for taxonomic recognition by an expert phycologist.

Image processing methodology

Images were processed according to the following flow of operations:

1) *Acquisition of microscope fields*

For image acquisition, 10 slides obtained by mixing samples from collection algae were processed. Fifteen microliters of sample were used for each slide (400 mm^2 coverslip surface, ≈ 30 µm sample height) and 1000 different microscope fields of each slide were acquired with adjusted white balance, taking as reference an empty portion of the slide. The coordinates of the bluish-gray color of this empty portion are stored as background color coordinates to be used in characteristic color determination, (operation 3). Boustrophedonic path was chosen for scanning.

It is very important to understand the fact that biological images are often far more difficult to be processed and recognized than daily-life images. Therefore, image acquisition is the most important step in image analysis since the goal is to achieve well-focused images with the lowest number of difficulties (or the highest information content) to tackle with in the successive processing. The microscope must be set at the best performance of Koehler illumination requirements following the indication of Zieler[33], the illumination must be even and uniform to avoid shadows, the flux emitted from tungsten lamp should be set so that the dark noise of the CCD camera has no influence, and the lamp color temperature should be set at about 3,000K for the best color balance, inserting gray filters in the optical path to avoid CCD camera saturation. The common habit to use digital operations in order to remove defects due to inaccurate acquisition should be avoided since these operations always reduce the information content of the image.

Debris, detritus, particles, bacteria, cell partly overlapping the slide border, empty dead cells, overlapping cells belonging to different algae, and out-of-focus cells are always present in a slide. In order to detect, identify, classify and enumerate algae with high taxonomic accuracy those objects must be detected and discarded in the successive digital operations.

2) Detection and recognition of algae and objects other than algae

The first operations acquired images undergo (Figures 1 and 2) are objects segmentation and contour detection.

Figure 1. Example of a microscope field at 40x.

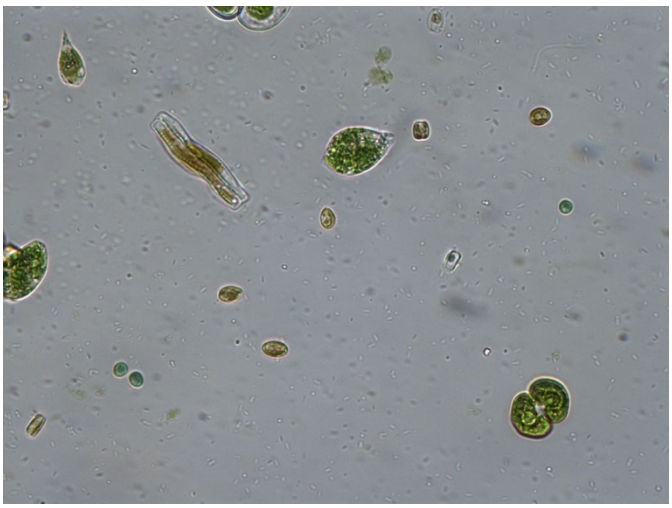

Figure 2. Example of a microscope field at 40x.

We define segmentation as the process of partitioning the images in a set of regions (i.e. the objects and the background) that collectively cover the entire image. We define a contour as a closed edge, i.e. the closed curve that delineates intensity transitions in the boundary between the object and the background. The algorithm calculates the points of the contours, the centroid distance spectrum, the dissimilarity measurement, and other morphological and densitometric features such as center of gravity coordinates, area, Feret diameters, extinction, etc. This algorithm used is an improved version of that already described in Coltelli *et al.* [32]. To obtain invariant features necessary for translation, rotation and scaling, the contour, if it exists, is normalized, oriented following the maximum Feret diameter, and uniformly resampled with 2n points. Ninety four numeric features (morphological, and densitometric) were actually calculated. To speed up algae grouping operation, only the features listed in Table 2 were used.

Table 2 List of morphological and densitometric features used in the processing

Feature	Description
Centroid distance spectrum	Distance of contour points from centroid
Coordinates of center of gravity	The coordinates of the center of gravity
Area	The number of square pixels inside the edge
Extinction	Integrated optical density
Dissimilarity measurement	Dissimilarity from circular shape
Size	Number of edges
Convex perimeter	Perimeter of the edge's convex hull
Moment Elongation	Ratio between the minimum and maximum moment of the edge (0÷1)
Feret maximum diameter	Measure of the maximum distance between the parallel lines tangent to an object size.
Feret minimum diameter	Measure of the minimum distance between the parallel lines tangent to an object size.
Feret Elongation	Ratio between the maximum and minimum Feret diameters.
Tortuosity	Ratio between the diagonal length of the bounding box and the length of the edges.
Circle Fit Radius	Radius of the circle that best fits the edge
Circle Fit Error	Average quadratic error of the fit.
Ellipse Fit Minor Axis	Minor axis of the ellipse that best fits the edges.
Ellipse Fit Major Axis	Major axis of the ellipse that best fits the edges.
Ellipse Fit Error	Average quadratic error of the fit.

The objects without a contour or with a contour with morphological and densitometric features not consistent with algae, such as debris, detritus, bacteria and particles were discarded. The objects partly overlapping the border of the acquired image were considered without a contour, hence discarded.

Algae connected in chains were identified as single unit when the chain possesses a distinguishable contour. Different shapes and appearance of the same cell, such as significantly different diatom girdle and valve views were identified as different cells. All the objects with features consistent with algae underwent the determination of characteristic color.

3) *Determination of algae characteristic color and removal of overlapping and out-of-focus algae*

After the conversion of the image colors from the RGB color space into the L*c*h* color space (Lightness, chroma, and hue), we calculated the occurrence of all the different colors of the recognized alga, in order to identify the characteristic color. The color histogram we obtained was fitted in a mixture of multivariate Gaussian distribution, using a maximum likelihood estimate of the component parameters. Each algal cell shows two regions: the chloroplasts characterized by an even distribution of the photosynthetic pigments and the cytoplasm which can be considered transparent as the background.

In the case of in-focus images, the mixture of multivariate Gaussian distribution shows two components, one corresponding to the color of the chloroplast compartment and the other corresponding to the color of the background. To assign each component to its own compartment, the coordinates of the colors corresponding to the means of the two components were compared with the coordinates of the color of the stored background (operation 1). The color having greater Euclidean distance corresponds to the chloroplast compartment and is defined the alga characteristic color, i.e. the color that represents the pigment signature of the taxonomic group the alga belongs to.

In case of out-of-focus cells, the mixture of multivariate Gaussian distribution shows three components, two corresponding to the colors of the chloroplast compartment and one corresponding to the color of the background.

Figure 3 shows the modifications of the mixture of multivariate Gaussian distributions induced by defocusing. The alga shown is the marine chlorophycean *Tetraselmis suecica*. Each alga image is accompanied by the corresponding bi- and three-dimensional visualization of the mixture of Gaussian distributions. The background color Gaussian component is not shown.

Out-of-focus cells together with objects having a contour but with an irregular color Gaussian distribution, such as empty cells, overlapping cells belonging to different algae, colored particles, etc. were discarded. The objects recognized as algae with their characteristic color and their morphological and densitometric features were organized in 53,869 vectors. This data set was divided into two subsets: 16,161 vectors (30%) were used to train the classification algorithm (training input data subset); while the remaining 37,708 vectors (70%) were set aside for validation and comparison of relative performances of the system in terms of taxonomic resolution.

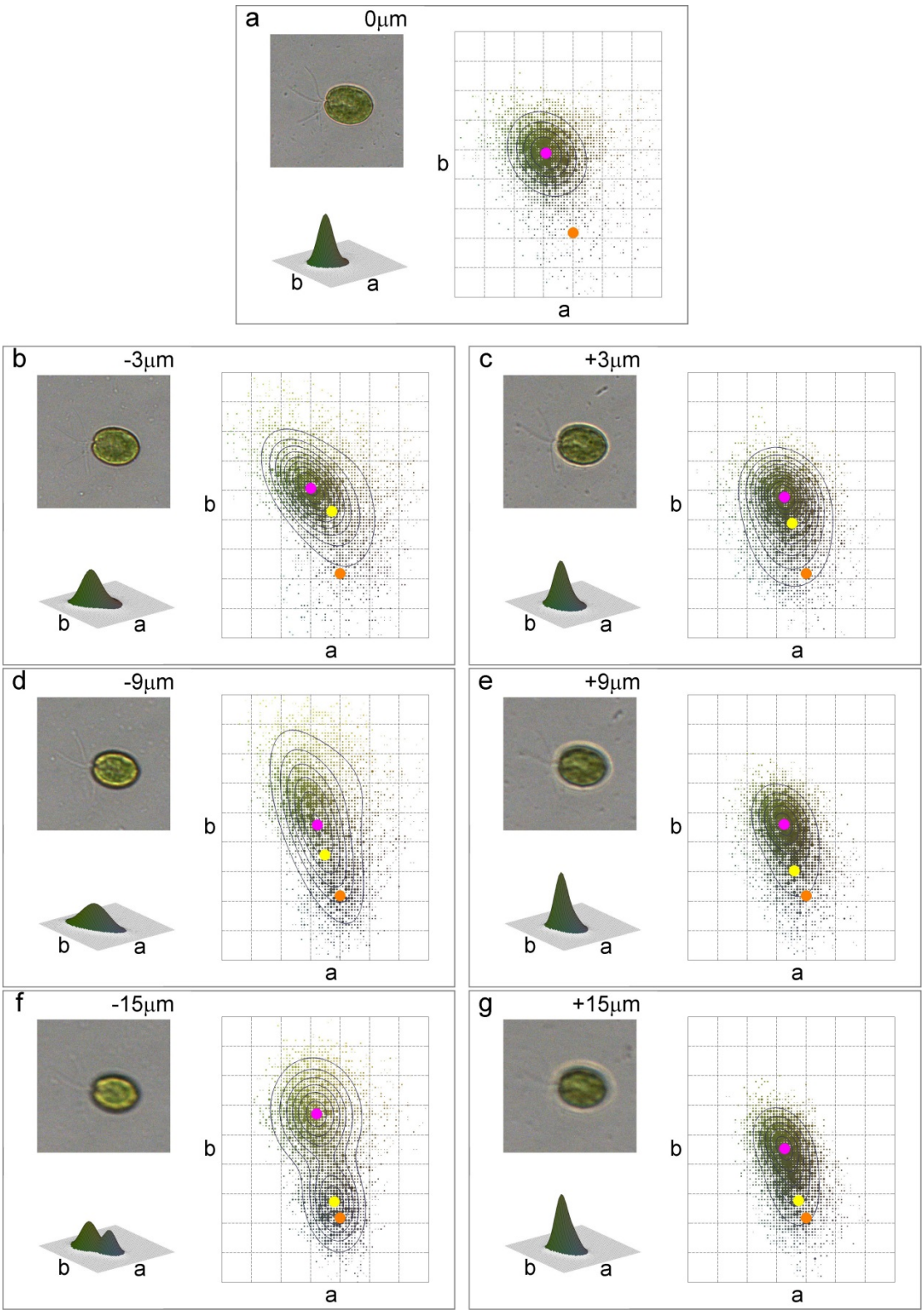

Figure 3 Modifications of the mixture of multivariate Gaussian distributions induced by defocusing. The alga shown is the marine chlorophycean *Tetraselmis suecica*. Figure 3a is the in-focus-image, while figures 3b-3g are the out-of-focus images obtaining moving the slide along the microscope z-plane (from -15 μm to +15 μm). Each alga image is accompanied by the corresponding bi- and three-dimensional visualization of the mixture of Gaussian distributions. In bi-dimensional visualization, the orange dot represents the coordinates of the stored background color, the magenta and yellow dots represent the mean color coordinates of the color components. The splitting of the color component becomes more and more evident increasing defocusing, (figures 3b-3g).

Journal Name RSCPublishing

ARTICLE

4) *Grouping and identification of algae*

To recognize algae, feature vectors should be grouped in classes. This grouping technique is called clustering. Feature vectors are the input data used to group algae by similarity by mean of an Unsupervised Neural Network. This network is a Self-Organizing Map (SOM) that consists of a two-dimensional structure of regularly spaced connected elements called neurons. The number of neurons is fixed as a round-off estimate of the number of taxonomic groups (genera and/or species). Each neuron is characterized by a feature vector of the same dimension of the input vector, with feature values randomly initialized. Neurons are connected to each other by neighborhood relations. The mathematical theory of the SOM is very complicated; for a more detailed explanation refer to Rissino and Lambert-Torres[34], Sap *et al.* [35], and Silva and Marques[36].

SOM algorithm is characterized by two steps: the training step that groups the algae according to the feature vectors, and the testing step (validation) that assigns algae images to the corresponding taxonomic group (genus and/or species).

In training, a sample input vector is drawn randomly from the training input data subset. This vector is fed to all the neurons in the map and a similarity measure based on Euclidean distance is calculated between the sample vector and all the neurons. The winning neuron, to which the sample vector is assigned, is the neuron with the highest similarity with it (or with the minimum Euclidean distance from it). After each feeding the vector of the winning neuron is updated to be a little closer to the input vector; the distances between the winning neuron and its neighbors are also similarly updated: the shorter the distance the closer the relation. The training ends when the map is no more modified by the sample data. Therefore, the SOM training step organizes the 16,161 sample vectors into homogeneous groups. Though reliable, these groups have no defined boundaries. Therefore, a segmentation procedure is necessary. Around each neuron an annulus is built having as radii two values of distance; within one map scan, the procedure connects the neighbor neurons which share input vectors in the overlapping annuli. The result of the segmentation procedure is a partitioned map whose regions represent the real taxonomic groups of algae (genera and/or species). At this step the number of cells belonging to each group is known, and therefore it is possible to calculate also the concentration of the different algae in the sample. Using the same procedure of the training, the validation step assigns each of the 37,708 features vectors (second subset) to its taxonomic group.

Results

Figures 1 and 2 show two different microscope fields acquired during operation 1. These fields simulate the content of a typical environmental sample: debris, detritus, bacteria, particles, algal cells partly overlapping the slide border, empty dead algae, overlapping cells belonging to different algae, and out-of-focus algae.

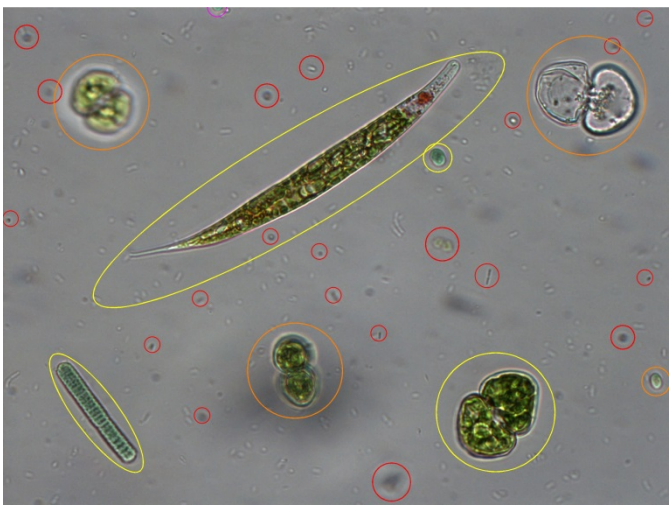

Figure 4. Output of operations 2, 3, and 4 on the image of figure 1. Objects identified as debris, detritus, bacteria and particles are red-framed; objects identified as algal cells partly overlapping the slide border are magenta-framed; objects identified as empty dead algae, overlapping cells belonging to different algae, and out-of-focus cells are orange-framed; objects identified as algae are yellow-framed.

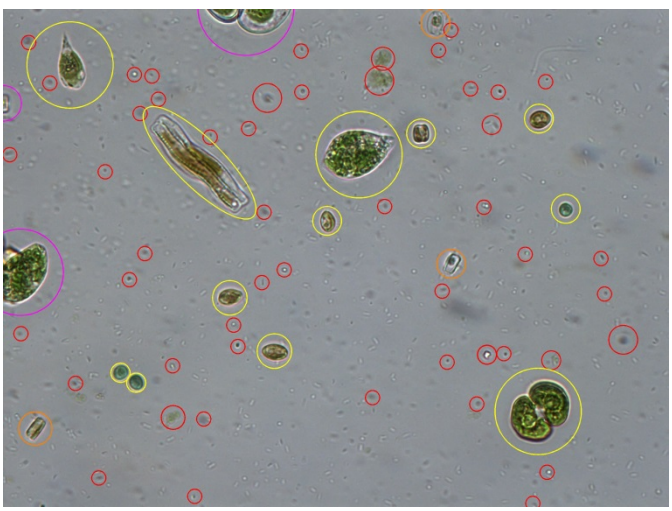

Figure 5. Output of operations 2, 3, and 4 on the image of figure 2. Refer to the legend of figure 4 for details.

Figures 4 and 5 show the output of operations 2, 3 and 4 on the images of figures 1 and 2. All the identified objects have been framed. Different frame colors correspond to different identification.

Figures 6 and 7 show some of the features extracted from the objects identified as algae in figures 4 and 5, such as invariant contour, centroid distance spectrum and dissimilarity measure, hue of the characteristic color, SOM group and taxonomic assignment validated by an expert phycology.

In these samples the dissimilarity measure, shown on the top right corner of the centroid distance spectrum, ranges from 1 (a perfectly round alga, such as *Cyanothece* sp. in figure 6a) to about 10^{20} (a long and very narrow alga, such as *Lyngbya* sp. in figure 6b). The color hue measure, shown on the top right corner of the color histogram, ranges from 11° (pinkish-red, such as *Haematococcus lacustris*) to 155° (blue-green, such as *Cyanophora paradoxa*); in samples shown in figures 1 and 2 it ranges from 87,3° (brownish green, such as the pennate diatom in figure 7f) to 145° (blue-green, such as *Cyanothece* sp. in figure 6a).

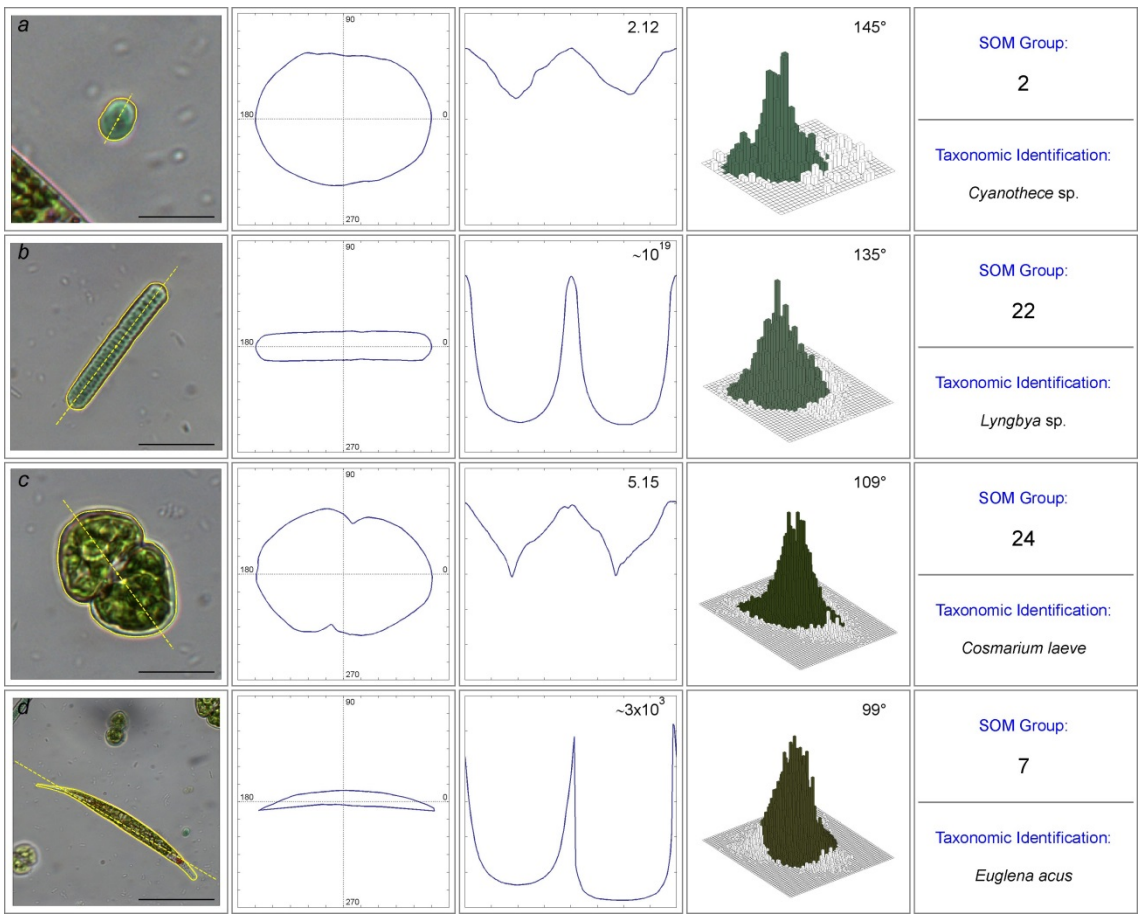

Figure 6. Features extracted from the objects identified as algae in figure 4. From left to right each row shows the segmented alga with its yellow contour and the maximum Feret diameter, (operation 2); the normalized and invariant contour, (operation 2); the centroid distance spectrum, with the dissimilarity measure (operation 2); the three-dimensional histogram showing only the chloroplast color component, with the color hue measure (operation 3); the alga SOM group (operation 4) and its taxonomic assignment (genus and/or species) after the expert validation.

Though the algae strains used in our test are 23 (Table 1), operation 4 resulted in 24 distinct groups because the centric diatom *Gomphonema parvulum* possesses significantly different valve (round) and girdle view (squared), which justify their identification as different cells. In the case of the pennate diatom only girdle views were identified.

Two different combinations of features were used to group the algae, (operation 4):

combination a) all the features of Table 2 (morphological and densitometric features) and the characteristic color (in *L*c*h** coordinates),

combination b) only dissimilarity measure and the characteristic color, (in *L*c*h** coordinates)..

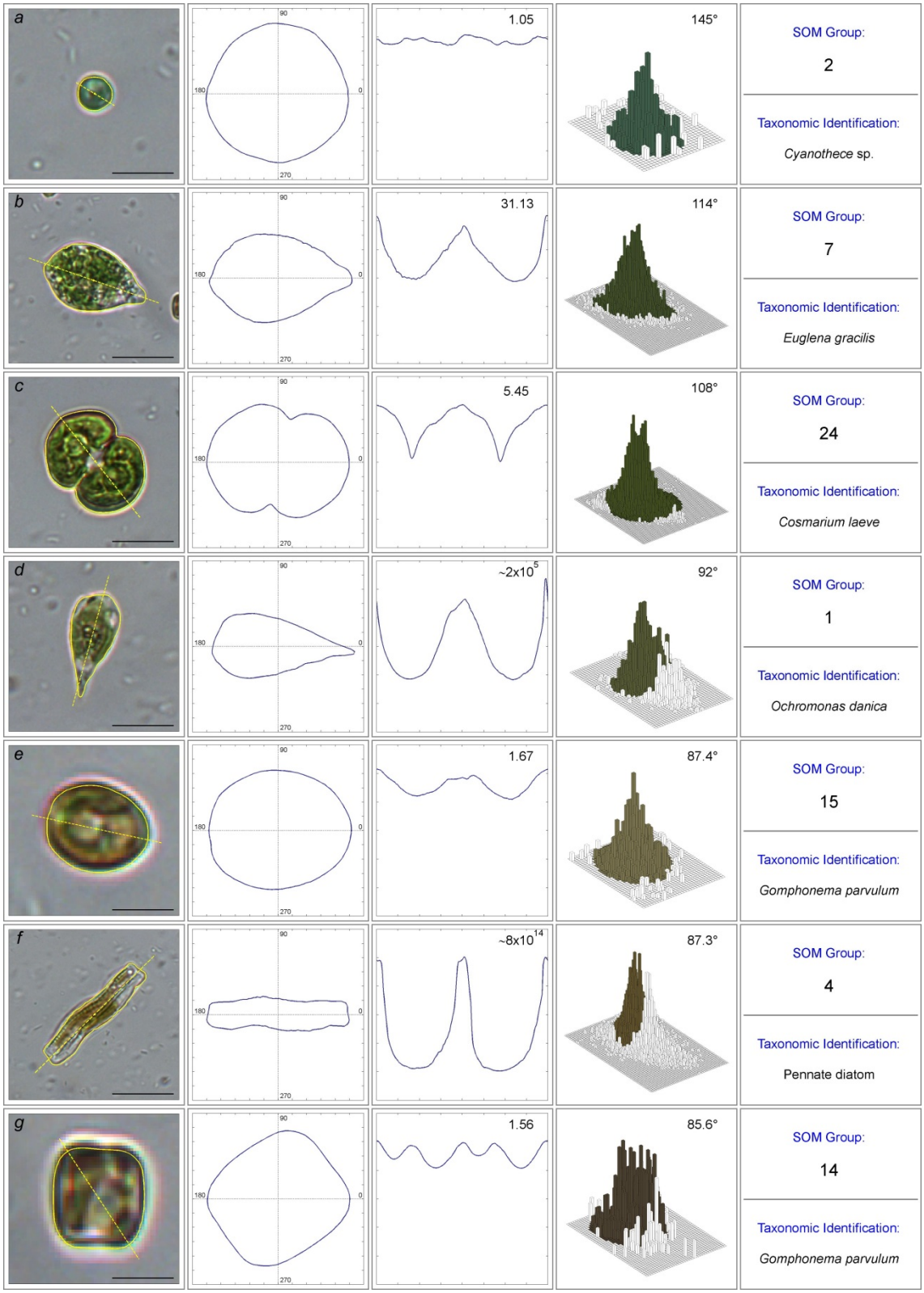

Figure 7. Features extracted from the objects identified as algae in figure 5; in case of multiple occurrences of the same alga only one representative is shown. Refer to the legend of figure 6 for details.

Figure 8a shows the result of operation 4 on the input data set using feature *combination b*: the 53,869 vectors were clustered and segmented into 24 homogeneous groups of algae with the highest taxonomic resolution (genus and/or species). Pink dots represent the algae used for training and validation; each of them corresponds to 20 feature vectors. Blue dots represent the

neurons, and the lines connecting them represent the feature distance between the groups. For sake of clarity, only two of the four dimensions (dissimilarity and hue), resulting from *combination b*, have been used for the representation. Notwithstanding, the 24 algal groups are clearly defined as shown in figure 8b. The numbers correspond to the SOM groups listed in Table 1.

The cell concentration calculated on the basis of the number of algae assigned to each group corresponded to the concentration of the algae present in the initial sample, (data not shown).

The result of operation 4 was verified by a phycology expert. The two combinations obtained taxonomic accuracy with no significant differences. The best accuracy, 98,6%, was reached by combination *a*, while combination *b* obtained 98,1% accuracy.

The time necessary for scanning a slide (1000 microscope fields) and building the input data set (operations 1, 2, and 3) is about 4.5 minute. Most of this time is spent in removing the out-of-focus-cells. The SOM training process takes about 3.5 minutes for an input data set of 1,000 feature vectors (combination *b*). Algal grouping by means of combination *a* takes about 5 minutes.

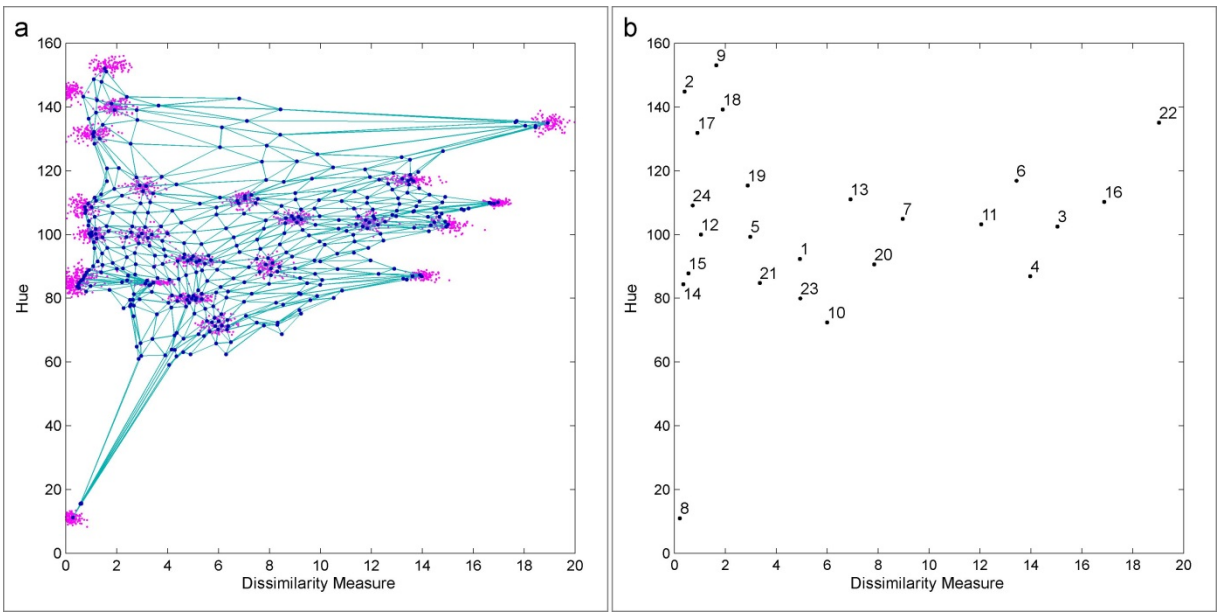

Figure 8. a: the result of operation 4 on the input data set; b: the 24 algal groups are clearly defined. The numbers correspond to the SOM groups listed in Table 1. Refer to text for details.

Discussion

The system we present is a valid tool to screen, identify and enumerate algal species with the finest taxonomic resolution; it obtains accuracy value higher than the values obtained by previously identification and classification methods, and almost equal to those obtained by a phycology expert, (about 99%).

This system marks a great step forward respect to the first basic and assisted version [32] of the methodology, approaching the final goal of our work, i.e. the field application for environmental monitoring.

It is based on training and validation databases containing a highly significant number of algal images (53,869). These images are acquired from slides that simulate real environmental samples in algae variety and concentration, selecting in-focus objects and discarding all the other possible causes of errors.

The higher accuracy value obtained is mainly due to the selection of the features necessary for the grouping operation, and to the implementation of a new, complex classification algorithm.

Different feature combinations were tested. The accuracies obtained using the features combination *a* and features combination *b* were the highest. They were not significantly different and very close to that of a phycology expert (about 99%), with only a slight difference in the processing time (combination *a* slower than combination *b*). Though other combinations with higher number of features (up to 94) were tested, the results were useless since they did not increase the accuracy and were too time consuming (data not shown). Our results demonstrate that the "characteristic color", i.e. the pigment signature of each alga, is a feature necessary for algal identification. It is perfectly suited to gather color information from algal samples and contains visual information equivalent to those that can be obtained by absorption microspectroscopy [37], with the advantage of an enormous simplification of the instrumental set-up. It should be stressed that the "characteristic color" is used exclusively by our system and not by the methods cited in the Introduction.

The new classification algorithm we implemented uses a complex unsupervised neural network, i.e a Self-Organizing Map, whose performance is enhanced by a joined segmentation procedure that refines the partitioned map and represents the regions of the real taxonomic groups of algae (genera and/or species) with the highest accuracy (Figure 8).

The only errors in the procedure are due to the misclassification of the girdle or valve view of the diatom *Gomphonema parvulum*, (Figure 7). While for the phycology expert both representations are easily recognized as belonging to the same alga, the automatic procedure mixed them up. However, these errors are formal since they do not have any practical consequences in the calculation of water quality assessment indices [38, 39].

The system will be completed by a piezoelectric controlled, continuous flow chamber, in order to allow in-continuous automatic analysis of environmental samples.

Conclusion

Algae assemblages have the potential to offer multifaceted characterization of water body conditions, and the stressor that may be affecting those conditions. It follows the importance of rapid and reliable bio-assessment based on qualitative and quantitative evaluation of algae assemblages. Bio-assessment allows the determination of baseline ecological conditions and/or the analysis of early warning signs of deteriorating conditions that can imply human health risks. Since the structure of the algae community differs from season to season, changes must be correlated to the seasonal baseline. Continuous systematic samplings of water bodies over an extended period of time are therefore necessary, which require the analysis of a huge amount of similar phytoplankton samples. Humans cannot deal with the volume of identification required when processing field samples from large scale surveys. Moreover, human analysis requires plenty of time, and fatigue and boredom produce severe loss of categorization performance in manual identification (> 50% error!). Therefore, the only feasible solution for water resources management is the use of automatic systems.

Our results show that water quality assessment in terms of both diversity and density of algal species can be performed by automatic methods. Notwithstanding, automatic methods still suffer from several defects, which should be eliminated before their commercialization. In our system at least 80 images of each group are required to train the SOM algorithm for a satisfactory performance; therefore, rare species can cause problems in grouping. Computation time depends first on the number of groups, secondly on the number of representative for each group, and thirdly on the number of features. For this reason it would be better to create different databases for each geographical location or ecological situation with precise temporal limits, in order to allow *in situ* monitoring.

References

1. J. Artiola, I. L. Pepper and M. L. Brusseau, *Environmental Monitoring and Characterization*, Elsevier Science & Technology Books, 2004.
2. G. B. Wiersma, *Environmental Monitoring*, CRC Press, Boca Raton, 2004.
3. B. Mitchell, *Resource and Environmental Management (2nd ed.)*, Pearson Education Limited, Harlow, 2002.
4. Weston, S. An overview of environmental monitoring and its significance in resource and environmental management, School of Resource and Environmental Studies, Dalhousie University. 2011.
5. *Water Monitoring*; FAO: Stockholm, 21 August 2006.
6. *Analytical Methods for Environmental Water Quality*; United Nations Environment Program Global, 2004.
7. U. N. Rai, S. Dubey, O. P. Shukla, S. Dwivedi and R. D Tripathi, *Environ. Monit. Assess.*, 2008, **144**, (1-3), 469-481.
8. L. Carvalho, C. McDonald, C. de Hoyos, U. Mischke, G. Phillips, G. Borics, S. Poikane, B. Skjelbred, A. L. Solheim, Van Wichelen, J. and A. C. Cardoso, *J. Appl. Ecol.*, 2013**, 50**, (2), 315-323.
9. *ACT Guidelines for Recreational Water Quality*, Health Protection Service. Australian Capital Territory, 2010.
10. U. B. Singh, A. S. Ahluwalia, C. Sharma, R. Jindal and R. K. Thakur, *Eco. Env. & Cons.*, 2013, **19**, (3), 793-800.
11. M. Mosleh, H. Manssor, S. Malek, P. Milow and A. Salleh, *BMC Bioinformatics*, 2012**, 13**, (Suppl 17), S25.
12. P. F. Culverhouse, R. Williams, M. Benfield, P. R. Flood,; A. F. Sell, M. G. Mazzocchi, I. Buttino and M. Sieracki, *Mar. Ecol. Prog. Ser.*, 2006**, 312**, 297-309.
13. R. Ellis, R. Simpson, P. F. Culverhouse and T. Parisini, *Neural Computing & Applications*, 1997**, 5,** (2), 99-105.
14. K. V. Embleton, C. E. Gibson and S. I. Heaney, *J. Plankton Res.*, 2003, **25**, (6), 669-681.
15. G. J. Kirkpatrick, D. F. Millie, M. A. Moline and O. Schofield, *Limnology and Oceanography*, 2000, **45**, (2), 467-471.
16. R. F. Walker, K. Ishikawa and M. Kumagai, *J. Microbiol. Methods*, 2002, **51**, (2), 149-162.
17. H. Xupeng, S. Rongguo, F. Zhang, X. Wang, H. Wang, and Z. Zheng, *Journal of Ocean University of China (Ocean Coastal Sea Res)*, 2010, **9**, 16-24.
18. E. Trampe, J. Kolbowski, U. Schreiber and M. Kühl, *Mar. Biol*,. 2011, **158**, (7), 1667-1675.
19. K. Furuya, M. Hayashi, Y. Yabushita and A. Ishikawa *Deep Sea Research Part II: Topical Studies in Oceanography*, 2003, **50**, (2), 367-387.
20. L. H. Zhang, J. Zhang and M. Cao, *Mar. Sci./Haiyang Kexue*, 2002, **26**, 60-65.
21. H. M. Sosik and R. J. Olson, *Limnol. Oceanogr. Methods*, 2007**, 5**, 204-216.
22. H. Mansoor, M. Sorayya, S. Aishah and A. A. M. Mogeeb, in *Automatic Recognition System for some cyanobacteria using image processing Techniques and ANN approach*, International Conference on Environmental and Computer Science, Singapore 2011, IACSIT Press, Singapore 2011.

23. S. B. Kamath, S. Chidambar, B. R. Brinda, M. A. Kumar, R. Sarada and G. A. Ravishankar, *Biosensors Bioelectron.*, 2005, **21**, (5), 768-773.
24. Z. Yao, M. Fei, K. Li, H. Kong and B. Zhao, *Neurocomputing*, 2007, **70**, (4–6), 641-647.
25. K. Rodenacker, B. Hense, U. Jütting and P. Gais, *Microsc. Res. Tech.*, 2006, **69**, (9), 708-720.
26. H. du Buf and M. M. Bayer, in *Automatic Diatom Identification*, eds. H. du Buf and M. M. Bayer, World Scientific Publishing Company, Singapore, 2002, Vol. 51, pp 289-298.
27. A. Verikas, A. Gelzinis, M. Bacauskiene, I. Olenina, S. Olenin and E. Vaiciukynas, *Expert Systems with Applications*, 2012, **39**, (5), 6069-6077.
28. A. Schaap, T. Rohrlack and Y. Bellouard, *Lab on a Chip*, 2012, **12**, (8), 1527-1532.
29. *Microscopic and Molecular Methods for Quantitative Phytoplankton Analysis*, Intergovernmental Oceanographic Commission, 2013.
30. N. Santhi, C. Pradeepa, P. Subashini and S. Kalaiselvi, *Bioinformatics and Biology Insights*, 2013, **7**, 327-334.
31. P. Coltelli and P. Gualtieri, *International Journal of Bio-Medical Computing*, 1990, **25**, (2-3), 169-176.
32. P. Coltelli, L. Barsanti, V. Evangelista, A. M. Frassanito, V. Passarelli and P. Gualtieri, *Environmental Science: Processes & Impacts*, 2013, **15**, (7), 1397-1410.
33. H. W. Zieler, *The optical performance of the light microscope Part.1*. Microscope Publications Ltd., London, 1972.
34. S. Rissino and G. Lambert-Torres, Rough Set Theory - Fundamental Concepts, Principals, Data Extraction, and Applications, Data Mining and Knowledge Discovery, in *Real Life Applications*, EDS. J. Ponce and A. Karahoca, InTec, 2009.
35. M. N. M. Sap and E. Mohebi, *International Journal of Signal Processing, Image Processing and Pattern Recognition* 2008, **1**, 11-20.
36. B. Silva and N. Marques, in *New Trends in Artificial Intelligence,* eds. J. Neves, M. F. Santos and J. Machado, Associacao Portuguesa Para a Inteligencia Artificial, Guimaraes, Portugal., 2007.
37. P. Gualtieri, *Crit. Rev. Plant Sci.* 1991, **9**, (6), 475-495.
38. A. E. Fetscher, R. Stancheva, J. P. Kociolek, R. Sheath, E. Stein, R. Mazor, P. Ode and L. Busse, *J. Appl. Phycol.*, 2014, **26**, (1), 433-450.
39. L. C. M. Palmer *Algae and water Pollution*, Municipal Environmental Research Laboratory, Cincinnati, Ohio, 1977.