Environmental Science
Processes & Impacts

RSC Publishing

PAPER

View Article Online
View Journal | View Issue



Cite this: *Environ. Sci.: Processes Impacts*, 2013, **15**, 1397

# Automatic and real time recognition of microalgae by means of pigment signature and shape

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

Received 25th March 2013
Accepted 29th April 2013

DOI: 10.1039/c3em00160a

**rsc.li/process-impacts**

Microalgae are unicellular photoautotrophic organisms that grow in any habitat such as fresh and salt water bodies, hot springs, ice, air, and in or on other organisms and substrates. Massive growth of microalgae may produce harmful effects on the marine and freshwater ecological environment and fishery resources. Therefore, rapid and accurate recognition and classification of microalgae is one of the most important issues in water resource management. In this paper, a new methodology for automatic and real time identification of microalgae by means of microscopy image analysis is presented. This methodology is based on segmentation, shape features extraction, and characteristic colour (*i.e.* pigment signature) determination. A classifier algorithm based on the minimum distance criterion was used for microalgae grouping according to the measured features. 96.6% accuracy from a set of 3423 images of 24 different microalgae representing the major algal phyla was achieved by this methodology.

**Environmental impact**

We present an innovative software methodology, which can offer a reliable, real time recognition of multialgal samples for environmental monitoring purposes. This software methodology combines robust image segmentation, shape features extraction, centroid distance spectrum calculations, and characteristic color (*i.e.* pigment signature) determination; it needs only a microscope equipped with a color CCD digital camera, and a personal computer. This methodology can achieve 96.6% accuracy, which has been tested using a set of 3423 images containing examples of 24 different microalgae representing the major algal phyla. At the present stage of development, the methodology cannot be considered ready for field analysis applications, mainly because of the lack of hardware automations, still it is very promising for future automated systems.

## 1 Introduction

Microalgae are unicellular photoautotrophic organisms that commonly grow in any habitat.[1] More than $10^6$ species have been recognized until now.[2] Microalgae form the base of the aquatic food chain; virtually all aquatic animals are dependent on these primary producers. Microalgae are also an excellent indicator of water quality (*i.e.* pollution), as their abundance and community composition most often reflects (and has the capacity to affect) the chemical and physical properties of water such as pH, nutrient levels, and transparency.[1] Their ability to grow in large quantities can lead to dramatic changes in the appearance, taste and odor of water, and can negatively affect organisms in higher trophic levels, because they may produce lethal neurotoxins (brain toxin) or hepatotoxins (liver toxin), which can cause serious illness or death in animals and humans in case of ingestion.[1] Microalgae can also interfere with general water usage and distribution, impairing water pumps, filters, pipes, animal troughs and misters, boilers and cooling equipment. Therefore, microalgae can be one of the most serious problems in aquatic ecosystem management.

In traditional research, microalgae investigation begins with direct sampling from the field. Back to the lab, researchers identify and classify the organisms present in the samples by optical microscopy. Microalgae show a wide variety of size, shape, texture and colors and all these characteristics are routinely used for taxonomical recognition.

However, this method is time-consuming and researchers must have abundant professional knowledge and experience of classification to achieve reliable results. In an attempt to speed up the time consuming procedure of classification by optical microscopy, several kinds of automatic analysis and microalgae identification methods were set up. They include for example methods based on algal cell morphology identification,[3] absorption spectroscopy,[4] fluorescence spectroscopy[5,6] high performance liquid chromatography[7,8] flow cytometry,[9] and gene probes.[10] Though automated classification of microalgae is a rather active field of research, commercial developments are rather limited.

A pioneering attempt in microalgal images automatic classification was that of Gorsky *et al.*,[11] who were able to distinguish between three species of distinct size and shape using simple geometric features. Culverhouse *et al.*[12] studied the classification

[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@pi.ibf.cnr.it; Fax: +39 050 315 2760; Tel: +39 050 3153026*

accuracy achieved by the neural network committee-based automated system DiCANN[13] and argued that the 72% accuracy achieved by the system in a six-class microalgal categorization task was similar to the accuracy achieved by trained personnel. Embleton *et al.*[14] trained a multilayer perceptron (MLP) to identify four species in lake water samples. Fourier descriptors, geometrical features, and histograms characterizing the grey level distribution of a region were used to select a set of features suitable for the training. Blaschko *et al.*[15] achieved 50% to 70% classification accuracy by categorizing microalgae into 12 classes plus an "unknown" class. To achieve this result, a large variety of features such as shape, moments, texture, contours (780 features in total) were used. Several classifiers, including decision trees, naive Bayes, ridge linear regression, neural network (NN), support virtual machine (SVM), as well as bagged and boosted ensembles were explored. Other works based on image processing with fuzzy logic, genetic algorithm, and neural network were used.[16–22] To capture more information for discrimination between five classes of microalgae, Rodenacker, *et al.*[23] applied fluorescence imaging in their image acquisition system. This approach was also used by Thar *et al.*,[24] Xupeng *et al.*[25] and Trampe *et al.*[26] to differentiate the abundance of different functional groups of microalgae and cyanobacteria in complex systems. Sosik and Olson[27] presented the most elaborated study regarding multiclass microalgal categorization using data obtained from Imaging FlowCytobot, which combine video and flow cytometric technologies. In total, 6600 visually identified and manually inspected images distributed across 22 categories were used in the study. In total, 210 features characterizing geometry, shape, symmetry, texture, and invariant moments of objects were extracted and 131 features were selected and used in an SVM classifier for the categorization. Cuiping, *et al.*[28] used a variety of moment features in an SVM with a radial basis function kernel and were able to distinguish between 241 species of marine phytoplankton with 89% accuracy.

Unfortunately, most of the existing research systems are adapted to identify only one or two specific types of algae. This is mainly due to the difficulties in choosing the correct feature pattern. For example, the Automatic Diatom Identification and Classification (ADIAC)[29] was set up to detect the differences between the frustules of centric diatoms only by using Fourier analysis of acquired images. ADIAC was effective but limited to the diatoms because they possess silica cell walls. Another technique, combining image segmentation, detection of circular objects, and stochastic optimization, was developed by Verikas *et al.*[30] in 2012. This study was limited to the dinoflagellate *Prorocentrum minimum*. The most recent paper in automated algae recognition and classification is that by Mosleh *et al.*;[31] the authors used only morphological parameters such as size and shapes and limited their study to five algal genera.

As alternative to the previously described techniques, often very complex and limited in their applicability, we present an innovative software methodology, which can offer a reliable, real time recognition of multialgal samples.

This software methodology combines robust image segmentation, shape features extraction, centroid distance spectrum calculations, and characteristic colour (*i.e.* pigment signature) determination. This methodology needs only a simple hardware platform, which integrates an optical microscope equipped with a colour CCD digital camera, and a personal computer. This methodology can achieve 96.6% accuracy, which has been tested using a set of 3423 images containing examples of 24 different microalgae representing the major algal phyla. A possible future application of this methodology is a commercial system for automated recognition and classification of different microalgae.

## 2 Materials

### 2.1 Algae cultures

Most samples belong to strain grown in our laboratory under controlled conditions,[1] but the system was tested also on samples taken from the field. Samples were collected from one of the eight main tributaries of the river Arno, in Tuscany. The water column was sampled at a depth of about 1 m at midflow from a road bridge, by means of a sampling bucket with a rope. Shallow areas at the side of the stream were also sampled by collecting the water. Samples were stored in plastic bottles and processed as soon as they reached the laboratory. Microscopic observations were performed on all the samples to check the composition of algal population. Table 1 shows the list of the algae used to test our system.

### 2.2 Operating platform for absorption microscopy

Fig. 1 shows the platform used to perform *in vivo* absorption microscopy measurements.

The hardware platform consists of a Zeiss Axioplan microscope (Zeiss, Oberkoken, DE), equipped with a 100 W tungsten

**Table 1** List of microalgae

| Name | Phylum | Source | Figure |
|---|---|---|---|
| *Closterium* sp. | Charophyta | Field | 8c |
| *Cryptomonas ovata* | Cryptophyta | Field | 7c |
| Cryptophyte | Cryptophyta | Field | 5m and 11d |
| *Cyanophora paradoxa* | Glaucophyta | Collection | 5l and 11c |
| *Euglena ehrenbergii* | Euglenozoa | Field | 8d |
| *Euglena gracilis* | Euglenozoa | Collection | 5k and 11b |
| *Gymnochlora stellata* | Cercozoa | Collection | 5i, 7d and 10e |
| *Gymnodinium acidotum* | Myzozoa | Field | 6d |
| *Haematococcus lacustris* | Chlorophyta | Collection | 6a and 9b |
| *Lyngbya* sp. | Cyanobacteria | Collection | 5n, 8e and 11e |
| *Mesostigma viride* | Charophyta | Collection | 5g and 10c |
| *Nannochloropsis* sp. | Ochrophyta | Collection | 5h and 10d |
| *Ochromonas danica* | Ochrophyta | Collection | 5f and 10b |
| *Pavlova lutheri* | Haptophyta | Collection | 5e, 6b and 10a |
| *Pediastrum duplex* | Chlorophyta | Field | 7a |
| Pennate diatom | Ochrophyta | Field | 5c, d and 9d, e |
| *Phacus acuminatus* | Euglenozoa | Collection | 6e |
| *Phaeodactlum tricornutum* | Ochrophyta | Collection | 8b |
| *Porphyridium purpureum* | Rhodophyta | Collection | 5a and 9a |
| *Prorocentrum micans* | Myzozoa | Collection | 5b, 7b and 9c |
| *Pyrocystis lunula* | Myzozoa | Collection | 7e |
| *Scenedesmus quadricauda* | Chlorophyta | Field | 5j and 11a |
| *Trachelomonas* sp. | Euglenozoa | Field | 6c |
| *Xanthonema* sp. | Ochrophyta | Collection | 8a |

 

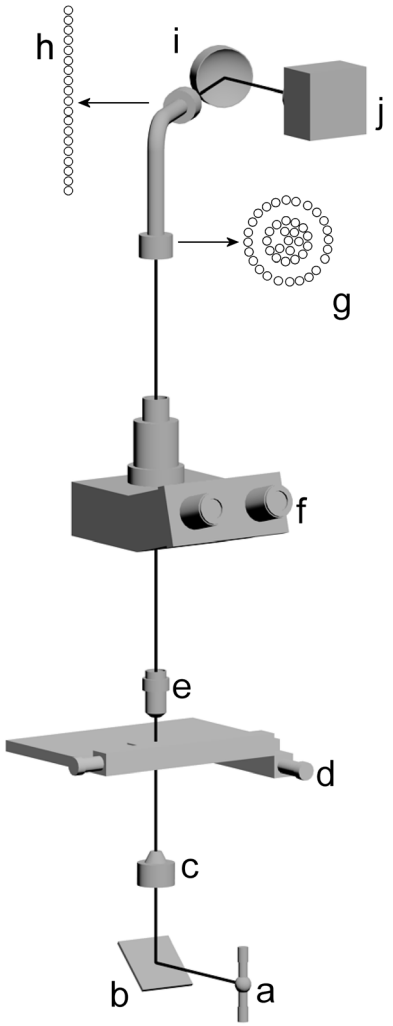


**Fig. 1** Schematic diagram of a simple microspectrophotometer: (a) lamp, (b) mirror, (c) condenser, (d) stage, (e) objective, (f) eyepiece, (g) light-guide probe entrance pupil, (h) light-guide probe exit pupil, (i) holographic grating monochromator, (j) slow scan cooled CCD camera.

lamp, an oil immersion condenser (N.A. 1.40), and 100× (N.A 1.3) and 40× (N.A. 0.75) plan apochromatic objectives. A high quality inspection probe (model 77423, Oriel, Stratford, Connecticut, USA) connected to a Schott KL 1500 probe illuminator system (Schott Corporation, Mainz, DE) was mounted on the back focal plane of the microscope. The entrance pupil of the probe consists of an outer bundle of 24 light-guides and a central bundle of 19 light-guides. The outer bundle is used for centering the objects, while the central bundle of the probe is used for acquiring the transmitted light. The 19 light-guides of the central bundle become vertically aligned at the level of the exit pupil of the probe, which is connected to a polychromator, a device in which the light is dispersed into its separate wavelengths by a diffraction grating (model 52300070, Jobin Yvon, Longjumenau, France). The flat dispersed image of the probe produced by the polychromator is in turn focused onto a digital slow scan cooled CCD camera (DTA Discovery DS260E, Pisa, Italy) equipped with an *IEEE 1394* interface and connected to a PC Intel core i7-2600 personal computer (Fig. 1).[32]

### 2.3 Operating platform for digital microscopy

Fig. 2 shows the platform used to perform *in vivo* and real time image measurements. The hardware platform consists of a Zeiss Axioplan microscope (Zeiss, FRG), with a 100 W stabilized tungsten-filament lamp, and a 100× (N.A. 1.3) plan apochromatic objective. A digital color CCD camera (Basler scA160028fm/fc, Basler Germany) equipped with an *IEEE 1394b* interface was mounted in the TV microscope path. The resolution of the original image is 1628 × 1236 pixels. The personal computer PC is Intel core i7-2600 (3.39 GHz), equipped with a 2 TB HD, 16 GB RAM, and Windows 7 operating system.

### 2.4 Image analysis

Image processing and pattern recognition were performed using MATLAB 7.0 software with home-made routines.[33]

## 3 Methods

### 3.1 Absorption spectrum determination

To perform the measurement the central bundle of the probe was centered on the photosynthetic compartment of the alga in such a way that one or two of the light guides were positioned outside the cell.[34]

Absorption measurements were based on the comparison of two radiant flux densities $I_S$ and $I_R$ (R stands for Reference, and S for Sample). $I_S$ results from the interaction of light with the sample (it is related to absorption cross-section of the molecules and the number of absorbing molecules), while $I_R$ results

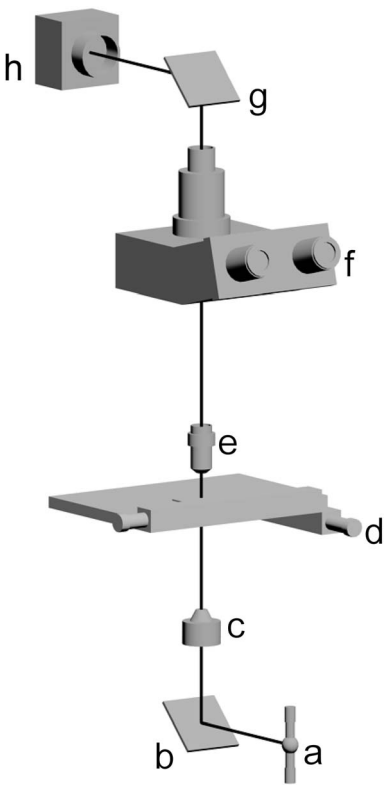


**Fig. 2** Schematic diagram of a simple microspectrophotometer: (a) lamp, (b) mirror, (c) condenser, (d) stage, (e) objective, (f) eyepiece, (g) mirror, (h) CCD color camera.

from the interaction of light with the reference material (outside the cell). Therefore, we can consider the absorbance of a sample ($A_S$) as derived from the measures as follows:

$$A_S = \log(I_R) - \log(I_S) \quad (1)$$

At 100× magnification, the diameter of the area being measured by each light guide was <1 μm. Measurements were performed on about 100 different cells per type, from 400 to 700 nm, with a resolution of 0.5 nm.

### 3.2 Image processing

Images were processed according to the following flow of operations:

(1) Image acquisition;
(2) Contours detection;
(3) Centroid distance spectrum calculation;
(4) Dissimilarity measure;
(5) Characteristic colour identification;
(6) Classification of images.

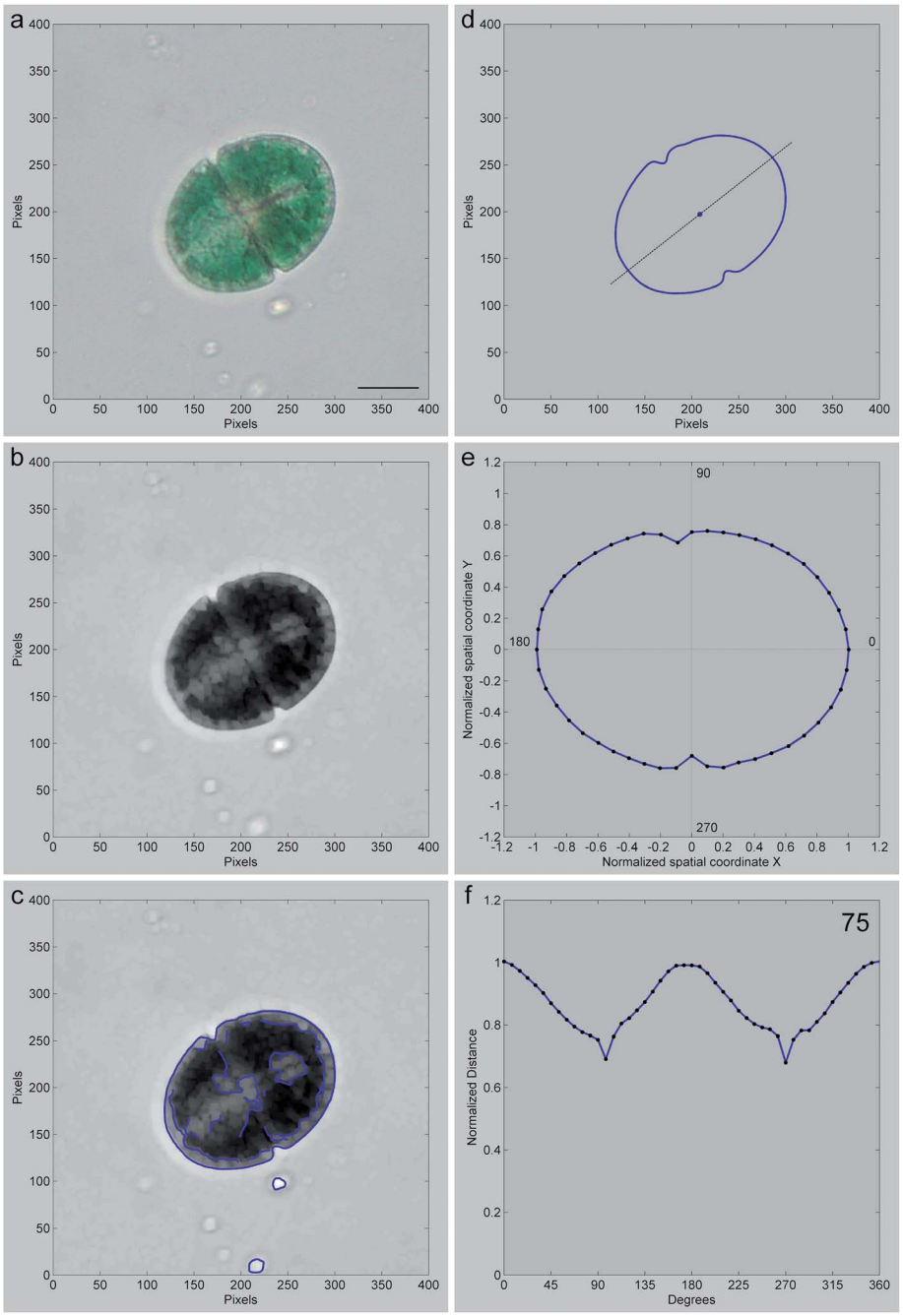


**Fig. 3** The algorithm for centroid distance spectrum calculation. See text for detail.

 

## 1 Image acquisition

For image acquisition, slides from both collection algae and field samples were prepared without any kind of processing or fixation. This is the most important step in image analysis, because image processing can remove unwanted information, not add new information. Therefore, all the color images were acquired with the highest accuracy. The microscope was set at the best performance of Koehler illumination requirements following the indication of Zieler[35] and the images were acquired with adjusted white balance, taking an empty portion of the slide as reference. The color of this reference represents the non-chloroplast domain used in Section 5c. Hundreds of microscopy field images were acquired resulting in a database of 3423 images of 24 different microalgae representing the major algal phyla (Table 1).

All the objects present in the microscopy field images underwent real time processing. Those corresponding to objects other than algae, such as debris, particles, detritus, overlapping cells, *etc.*, which did not possess a contour (see Section 2) and/or a histogram with two Gaussian components (see Section 5) were discarded and not added to the database.

Cells connected in chains are recognized as a single unit when they possess a distinguishable contour, (Fig. 8a); when cells are in stacks consisting of non-distinguishable cells, *i.e.* only the stack possesses a contour, the stack itself is recognized as an alga object, (Fig. 11a). In the case of different shapes of the same algal cells, these shapes are assigned to different database classes.

We used two different microalgae to show the results of the different operations, *i.e.* a dinoflagellate (Fig. 3a) for operations 1–4, and a glaucophyte (Fig. 4a) for operations 5–6. This choice was made only for aesthetic purposes. Routinely, the whole flow of operation is applied to each microalgal image.

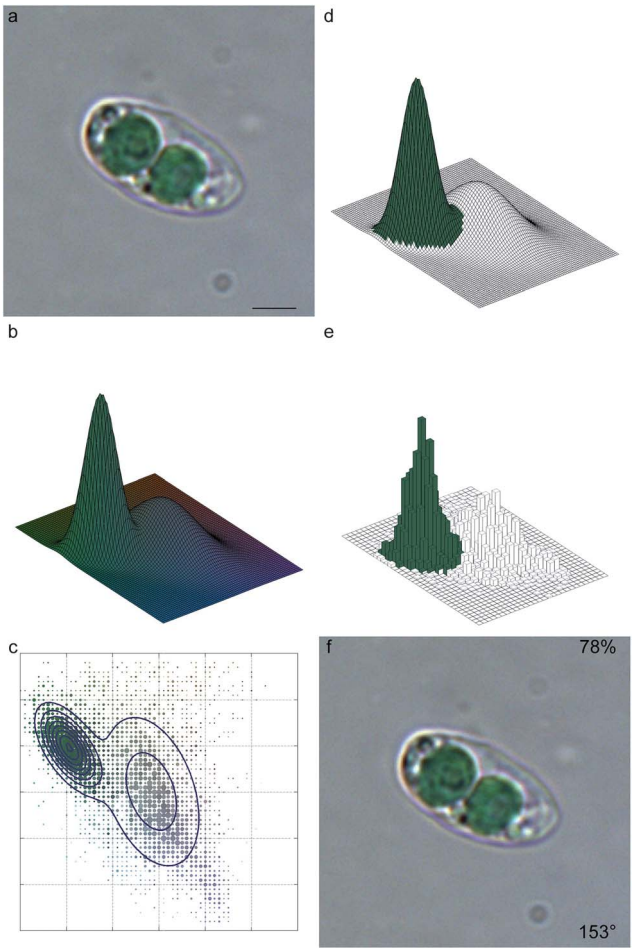


**Fig. 4** The algorithm for characteristic colour calculation. In frame (c), the dots sizes are proportional to the colour occurrences. See text for other details.

## 2 Contours detection

The amount of data of the acquired images was reduced by representing the algae with lengths and contours as descriptors. We define a contour as the closed edge, the edge being the curve that delineates intensity transitions in the boundary between the object and its background.

Finding contours consisted of several steps:

(2a) The acquired color image, in the red, blue and green color space (RGB), (Fig. 3a), was converted into a gray scale image. The following standard conversion equation directly computed a gamma-corrected luma, *i.e.* $Y'$ or electronic brightness (voltage of display), which is used as black and white information:

$$Y' = 0.299\mathrm{R} + 0.587\mathrm{G} + 0.144\mathrm{B} \quad (2)$$

(2b) Several pre-processing operations were performed on the gray scale image such as contrast enhancement, systematic errors compensation, and morphological operations (*i.e.* close operation), (Fig. 3b).[36]

(2c) The next step was a filtering process that provided an enhanced gray scale image of the edges calculating gradient magnitude at each pixel in the image.

$$\text{Gradient Magnitude} = \sqrt{{I_x}^2 + {I_y}^2} \quad (3)$$

where $I_x$ and $I_y$ are the $x$ and $y$ derivative values respectively, which define the components of the gradient vector as:

$$\text{Gradient vector} = \begin{bmatrix} I_x \\ I_y \end{bmatrix} \quad (4)$$

The stronger the intensity transition, the greater the magnitude.

(2d) The fourth step consisted of the detection and thresholding operations performed on the enhanced gradient magnitude gray scale image that determines the edgels (elementary points within an edge), and their $X$ and $Y$ coordinates. Any edgel is located at the maximum value of the gradient magnitude over adjacent pixels, in the direction defined by the gradient vector. At the end of this procedure, a black and white image of edgels is obtained.

(2e) The final step 8-connected the neighboring edgels to build an edge chain, respecting that: consecutive edgels must occupy separate 8-connected pixels, no branches are allowed (*i.e.* they start or end a separate chain), and no edge wider than one pixel is present, (*i.e.* no two-by-two blocks of pixels containing edgels in the edge chain can exist). Features were

calculated for each edge. At the end of this procedure home-made routines were applied to control the degree of smoothness of the edges, and to link the extremities of non-closed edges together to form a contour filling short gaps. In Fig. 3c the edge layer was superimposed on the gray scale image. Based on calculated features such as the average steepness of brightness variations, size and closure, the microalga contour was selected (Fig. 3d). Morphological microalgal features such as center of gravity coordinates, Feret diameters size, skewness and kurtosis were also calculated.

## 3 Centroid distance spectrum calculation

To obtain invariant features for translation, rotation and scaling, the contour is normalized, oriented following the maximum Feret diameter, and its points uniformly resampled (Fig. 3e). Using this normalized contour a centroid distance ($CD_i$) function was calculated with the following formula:

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

where $x_c$ and $y_c$ are the centroid coordinates and $x_i$ and $y_i$ are the coordinates of the points of the contour. In this study, we used 360 points, one for every degree. In Fig. 3f the Centroid Distance Spectrum (CDS) is shown.

## 4 Dissimilarity measure

To reduce the number of features in the classification we used the Dissimilarity Measure Criterion (DMC). It represents the similarity relation between the alga contour vector and the circumscribed circumference vector. These two vectors have the same number of points.

$$DMC = 1 - \frac{C_{ma} \bullet C_c}{\|C_{ma}\| \cdot \|C_c\|} = 1 - \frac{\sum_i C_{ma_i} \cdot C_{c_i}}{\sqrt{\sum_i {C_{ma_i}}^2 \cdot \sum_i {C_{c_i}}^2}} \quad (6)$$

where $C_c$ is the circumference contour vector, $C_{ma}$ is the alga contour vector, and the big dot is the scalar product between the two vectors.

As a result:

coincidence $= 0 \leq DMC \leq 2 =$ maximum dissimilarity.

For a greater variation range the following scale is used:

$$DMC_{exp} = 10^{100 \cdot DMC} \quad (7)$$

In this case:

coincidence $= 1 \leq DMC_{exp} \leq 10^{200} =$ maximum dissimilarity.

The dissimilarity measure is shown in the top right corner of Fig. 3f.

## 5 Characteristic colour identification

(5a) The first step was to convert the RGB color model in the $L^*a^*b^*$ color space. The three coordinates $L^*$, $a^*$, and $b^*$ represent respectively the lightness of the color ($L^* = 0$ means black and $L^* = 100$ means diffuse white), the position between red/magenta and green ($a^*$ negative values indicate green while positive values indicate magenta) and the position between yellow and blue ($b^*$ negative values indicate blue and positive values indicate yellow). The purpose of $L^*a^*b^*$ color space is to work with a space that is perceptually uniform and whose coordinates can be easily calculated from the RGB color model. Perceptually uniform means that a change of a color value should produce a change of the same visual importance, *i.e.* the Euclidean distance in the three-dimensional space between colors can be used as colors perceptual distance. The most important attribute of the $L^*a^*b^*$ color space is the device independency. This means that the colors are defined independently of their nature of creation or the device that is used to display them. Another attribute of the $L^*a^*b^*$ color space is that its color gamut exceeds that of the RGB color model.

The $L^*a^*b^*$ color space is then transformed into $L^*C^*h^*$ color space that uses cylindrical coordinates. $L^*$, $C^*$, and $h^*$ correspond to the perceptual attributes of lightness, chroma and hue. $L^*$ is the lightness coordinate as in $L^*a^*b^*$; $C^*$ is the chroma coordinate, which measure the distance from the lightness axis; and $h^*$ is the hue angle expressed in degrees with 0° being a location on the $+a^*$ axis, continuing to 90° for the $+b^*$ axis, 180° for $-a^*$, 270° for $-b^*$, and back to 360°. Mathematically, chroma ($C^*$) and hue angle ($h^*$) are defined as:

$$C^* = \sqrt{(a^*)^2 + (b^*)^2}$$
$$h^* = \arctan\left(\frac{b^*}{a^*}\right) \quad (8)$$

For the complete list of conversion formulae between color spaces, refer to Ford and Roberts.[37]

(5b) After the conversion in the $L^*C^*h^*$ color space, we calculated the occurrences of all the different colors of the segmented cell, in order to identify the characteristic color of the microalgae (Fig. 4a). The color histogram we obtained was fitted in a mixture of multivariate Gaussian distributions, defined by their means and covariances, by means of the Matlab gm distribution fit function. This function uses a maximum likelihood estimate of the component parameters. In our case the mixture has two components, *i.e.* the chloroplast domain and the non-chloroplast domain. Fig. 4b shows the three-dimensional Gaussian fitted color histogram on the $a^*$, $b^*$ plane while Fig. 4c shows the isolines projection of the same histogram on the same $a^*$, $b^*$ plane. In both cases an average lightness was used.

(5c) Successively, the coordinates $L^*$, $C^*$, and $h^*$ of the colors corresponding to the means of the two Gaussian components were compared with the coordinates of the colors of the non-chloroplast domain of the reference image. The color having greater Euclidean distance corresponds to the chloroplast domain and is defined as the alga characteristic color, *i.e.* the color that represents the pigment signature of each algal group. To demonstrate the effectiveness of this result at a visual inspection, we added two further steps to the process. In Fig. 4d all the colors belonging to the chloroplast domain in the Gaussian fitted color histogram have been replaced by the characteristic color. Fig. 4e shows the result of the same replacement in the calculated original histogram. In Fig. 4f the characteristic color replaces the colors of the chloroplast in the

 

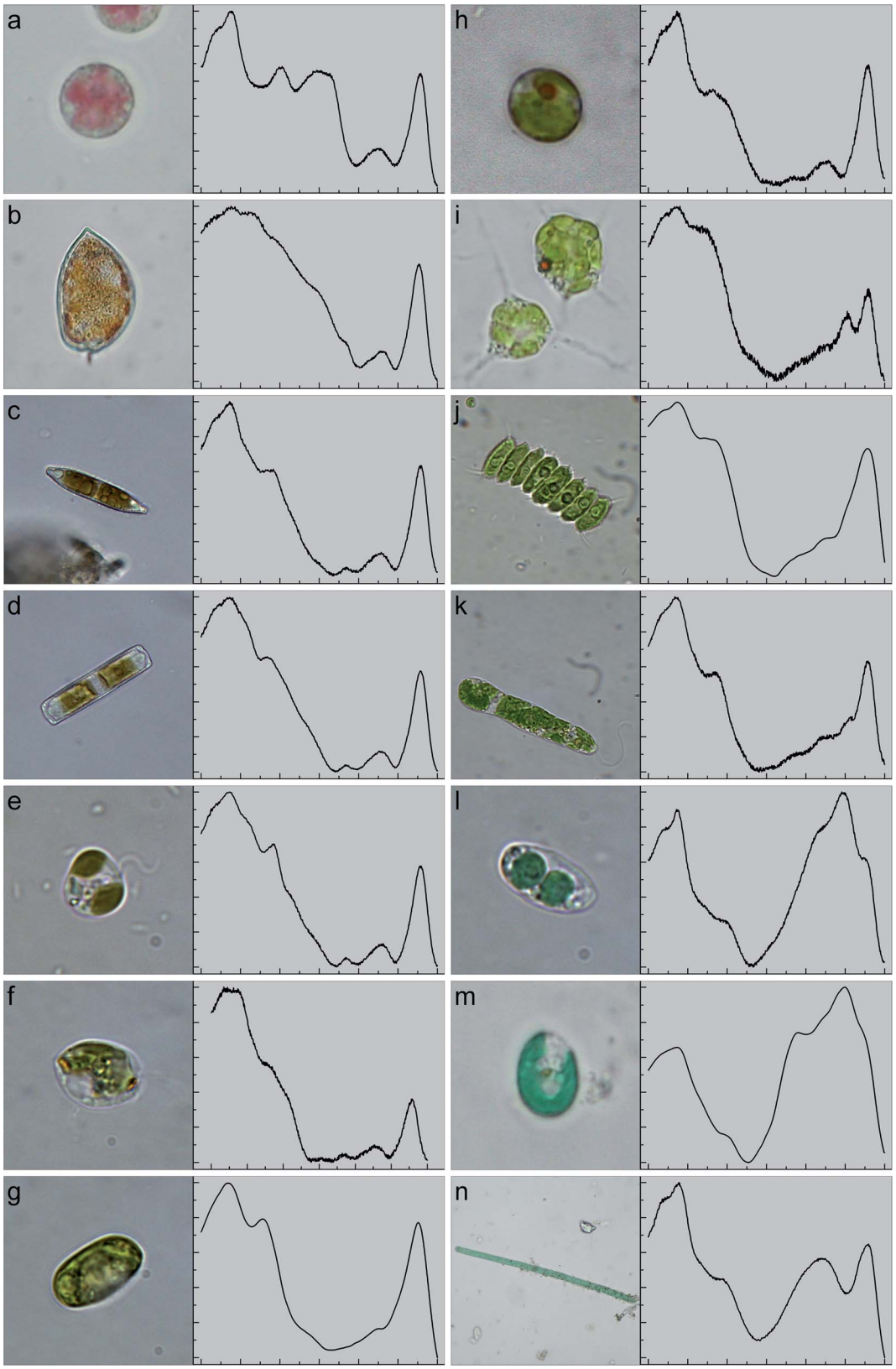


**Fig. 5** Results of microspectrophotometric measurement on photosynthetic compartments; each series shows the sample used and the absorption spectrum.

original image shown in Fig. 4a: the chromas and hues of the original colors are replaced with the chromas and hues ($C^*$, and $h^*$) of the characteristic color; only the lightness coordinate ($L^*$) varies according to the value of the original image. The number in the bottom right corner represents the characteristic color hue ($h^*$), while the number in the top right corner represents

the percentage of color substitution. The percentage of pixels substituted in this cell is about 78. The number of colors in the cell decreases from 9255 to 192.

## 6 Classification of images

All the features measured by the operations that have been described so far (*i.e.* absorption spectrum, centroid distance spectrum, dissimilarity measure, and characteristic color) are used to classify the microalgae according to the minimum distance criterion. Each class of objects (corresponding to an algal grouping) is represented by a prototype pattern vector. The Minimum Distance Classifier computes the Euclidean distance between the vector of the features of the microalga ($\boldsymbol{x}$) to be classified and each of the prototype vectors present in the database ($\boldsymbol{m}_j$). The microalga is assigned to a class based on the smallest distance. This distance $D_j$ is calculated as:

$$\mathrm{D}_j(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{m}_j\| \; j = 1,2,\ldots,W \tag{9}$$

where $\|\boldsymbol{a}\| = (\boldsymbol{a}^T\boldsymbol{a})^{1/2}$ is the Euclidean norm of a generic vector $\boldsymbol{a}$, $W$ is the number of classes, and $\boldsymbol{m}_j$ is the mean vector of the patterns of a class (or prototype), *i.e.*

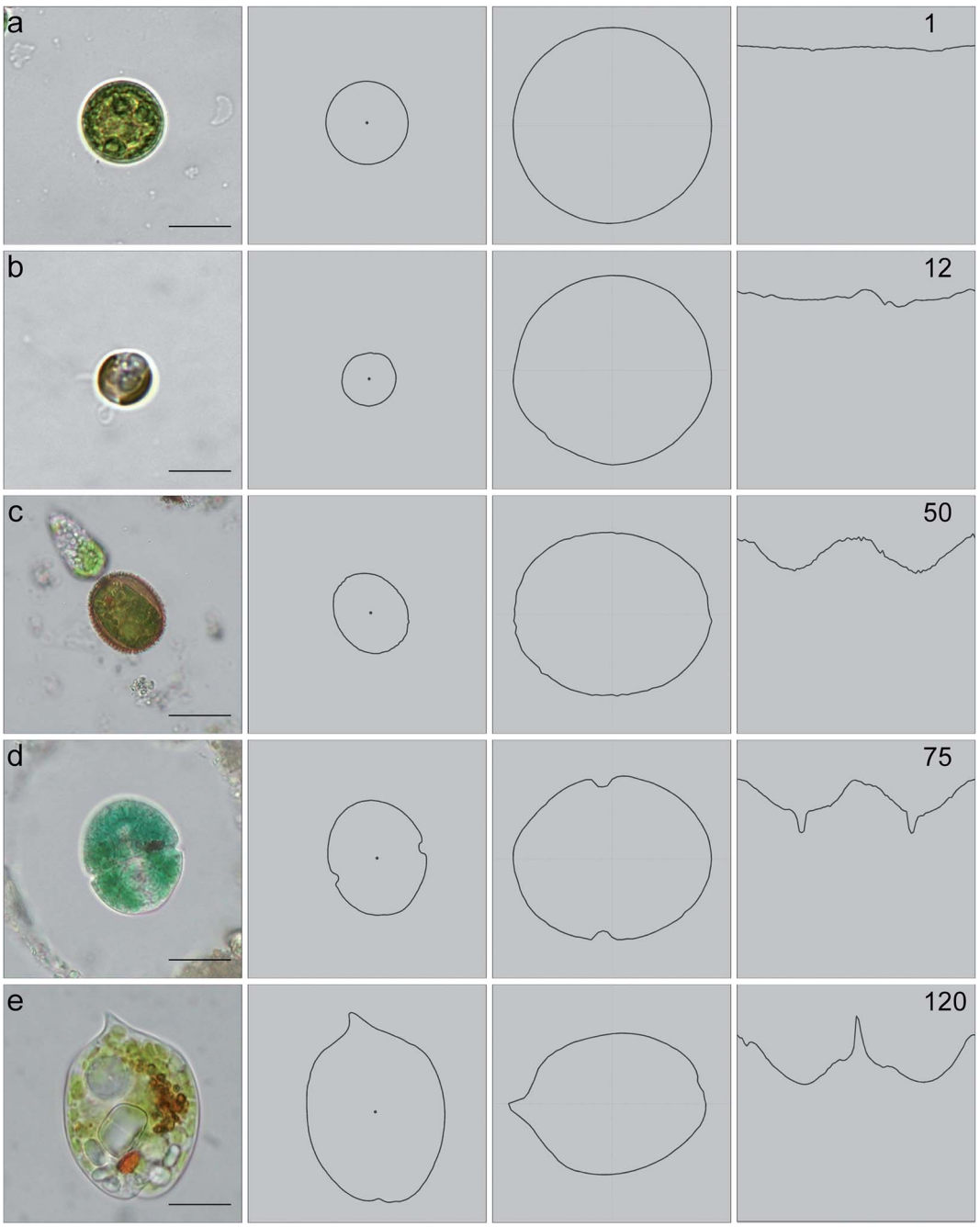


**Fig. 6** List of microalgae ordered by their DMC. Range from 1 to 120.

 

$$\boldsymbol{m}_j = \frac{1}{N_j} \sum_{\boldsymbol{x} \in \omega_j} \boldsymbol{x}_j \quad j = 1, 2, \ldots, W \tag{10}$$

If $D_j(\boldsymbol{x})$ is the smallest distance, we then assign $\boldsymbol{x}$ to the class $\omega_j$.

The decision boundary surface between two classes $\omega_i$ and $\omega_j$ is the perpendicular bisector (hyper plane) of the line segment joining $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$.

The image database contains 3423 images of 24 different microalgae representing the major algal phyla, hence in our study, $W$ is equal to 24.

## 4 Results

Photosynthetic compartments belong to both algae and plants. Absorption spectra measured *in vivo* by microspectrophotometry on these compartments can give us very precise and accurate information about the spectral range in which pigment molecules capture photons. Pigment distribution is almost constant in plants, while each alga is characterized by its own absorption spectrum, which represents the pigment signature of the alga. Fig. 5 shows in detail the different absorption spectra of the algae measured with the microspectrophotometric set-up

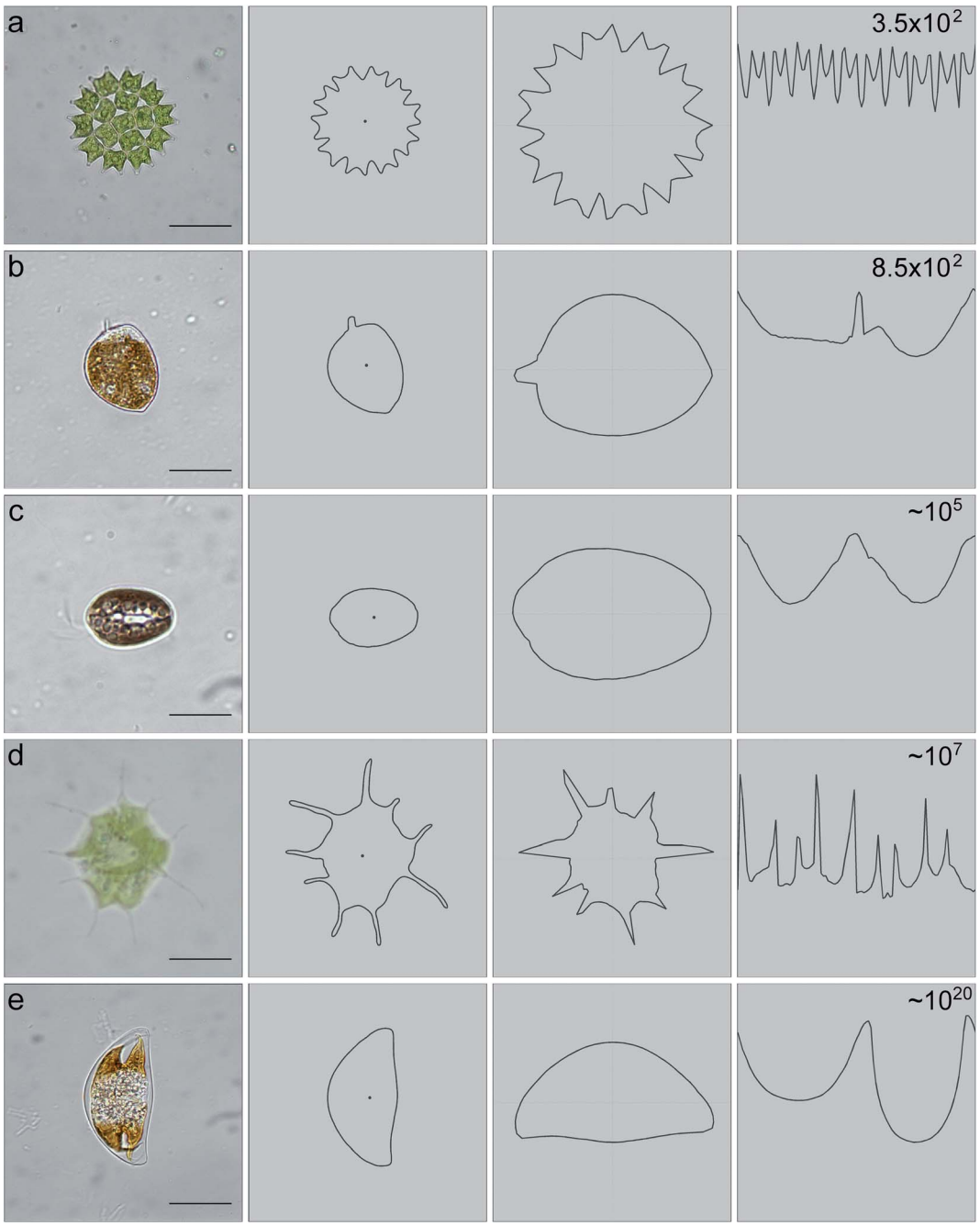


**Fig. 7** List of microalgae ordered by their DMC. Range from $3.5 \times 10^2$ to $10^{20}$.

described above (Fig. 1). We measured the absorption spectra in algae cultivated under laboratory conditions. In natural ecosystems, factors such as light quality and quantity, nutrient availability, salinity and temperature variations, and growth phase may affect pigmentation dynamics.

Fig. 6, 7 and 8 show in detail respectively from left to right: the original image, its contour, its normalized and invariant contour, and the centroid distance spectrum with the corresponding dissimilarity measure. Microalgae images were acquired with the digital microscope set-up described above (Fig. 2). The centroid distance spectrum varies from a horizontal line, *i.e.* all the points are coincident with the circumscribed circumference, to a curve that has only three points of coincidence, as expected from elementary geometry. The dissimilarity measure, shown on the top left corner of the centroid distance spectrum images, ranges from 1 corresponding to a perfect circular microalga to $10^{150}$ corresponding to a long and very narrow microalga. These values, in ascending order, are used to list the selected images.

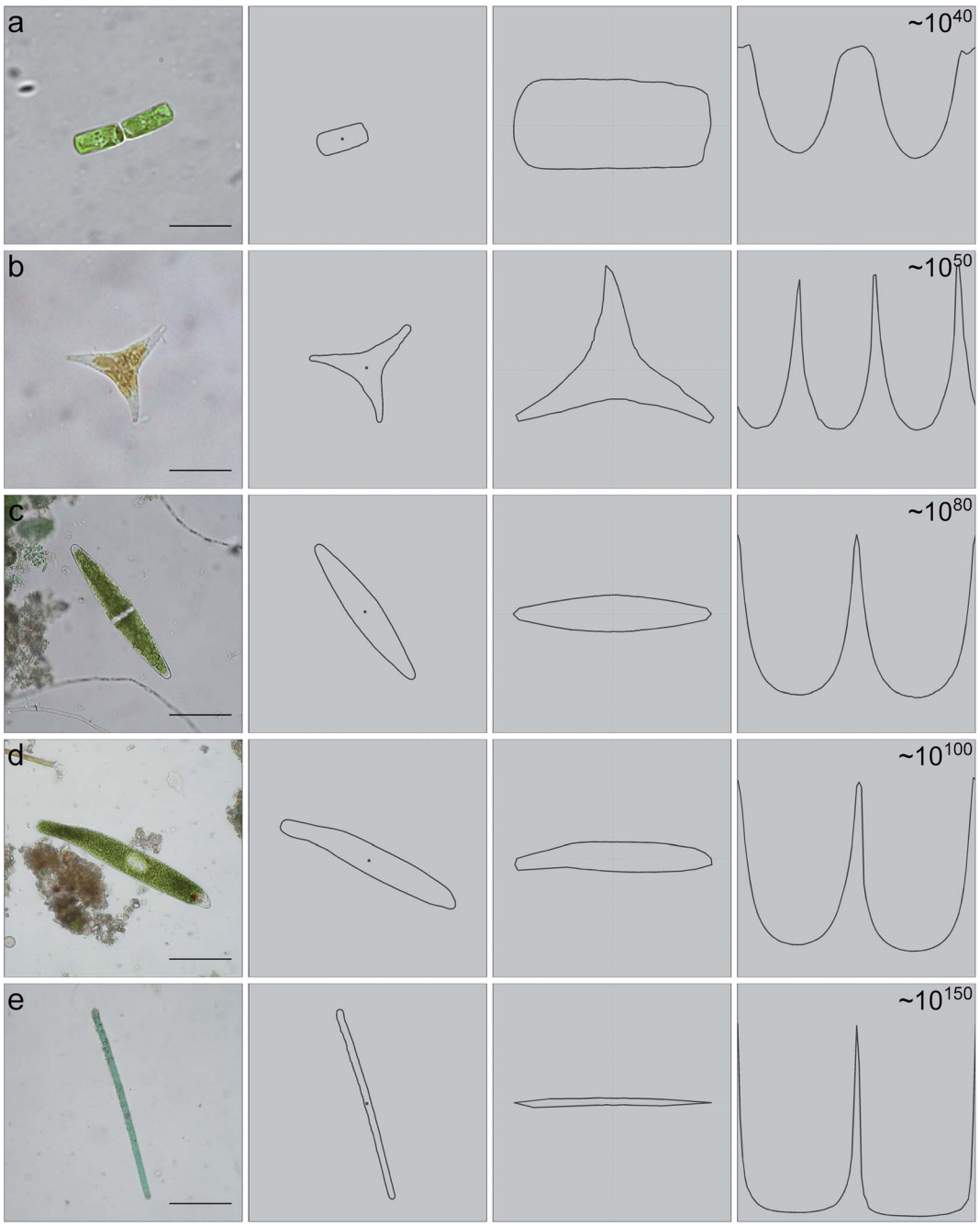


**Fig. 8** List of microalgae ordered by their DMC. Range from $10^{40}$ to $10^{150}$.

 

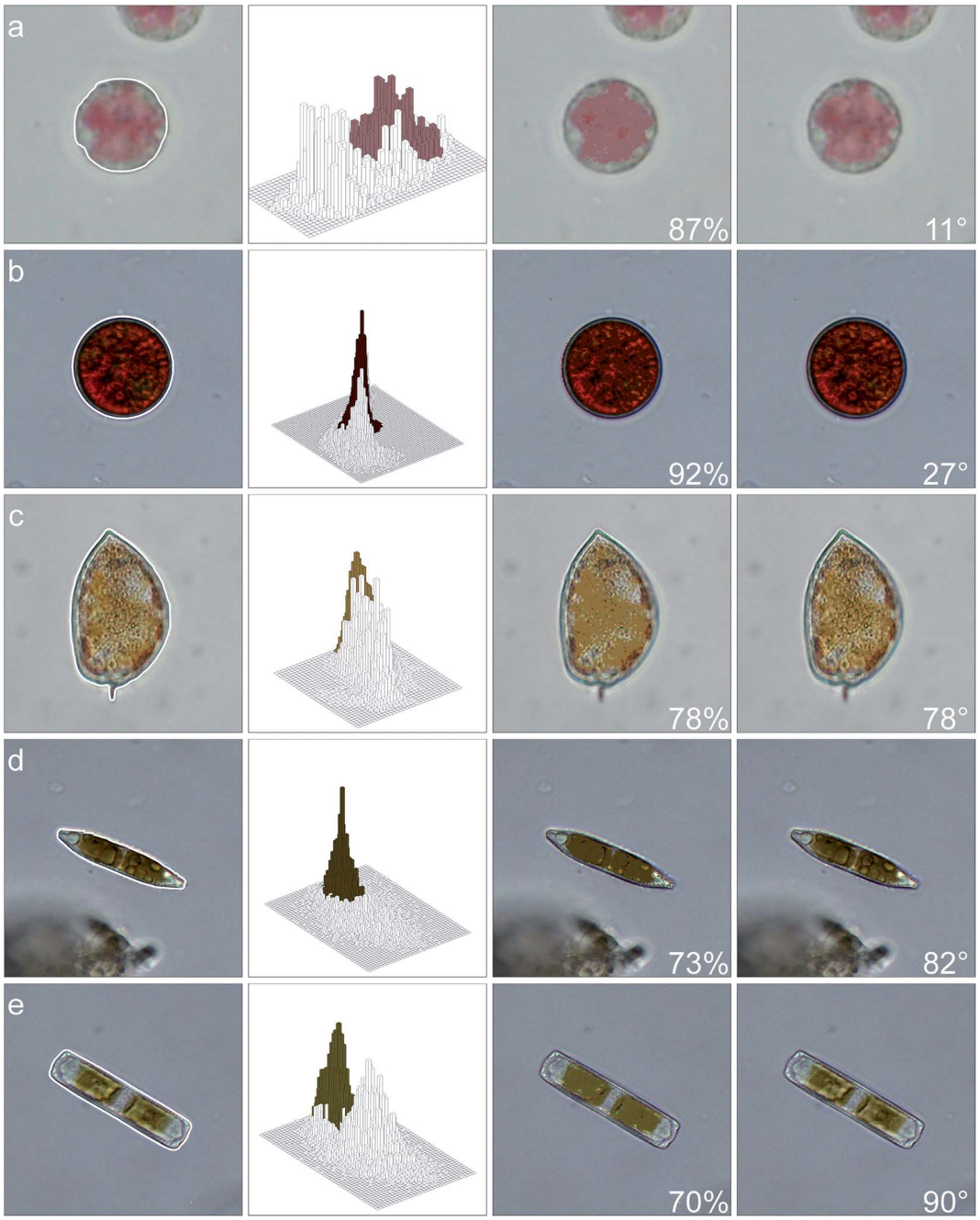


**Fig. 9** List of microalgae ordered by characteristic color hue (CCH). Range from 11° to 90°.

Fig. 9, 10 and 11 show in detail respectively from left to right: the original image with the contour drawn as white overlay, the calculated original histogram after characteristic color replacement, and the original image after characteristic color replacement. In the last frame the original image after characteristic color replacement is shown by varying the lightness coordinate ($L^*$) according to the original image. The perceptive differences between the original image (first frame) and the original image after characteristic color replacement and lightness variation (fourth frame) are undetectable. The image in the third frame is shown to appreciate the percentage of pixels involved in color replacement. The corresponding value is shown in the bottom right corner. The hue of the characteristic color is shown in the bottom right corner of the fourth frame. It ranges from 11° to 164°, *i.e.* from a pinkish-red to blue-green. This is the color scale of the algae.

## 5 Discussion

A total of 3423 images containing examples of 24 microalgae representative of the major algal phyla were processed and the correct number of database classes and the correct assignment

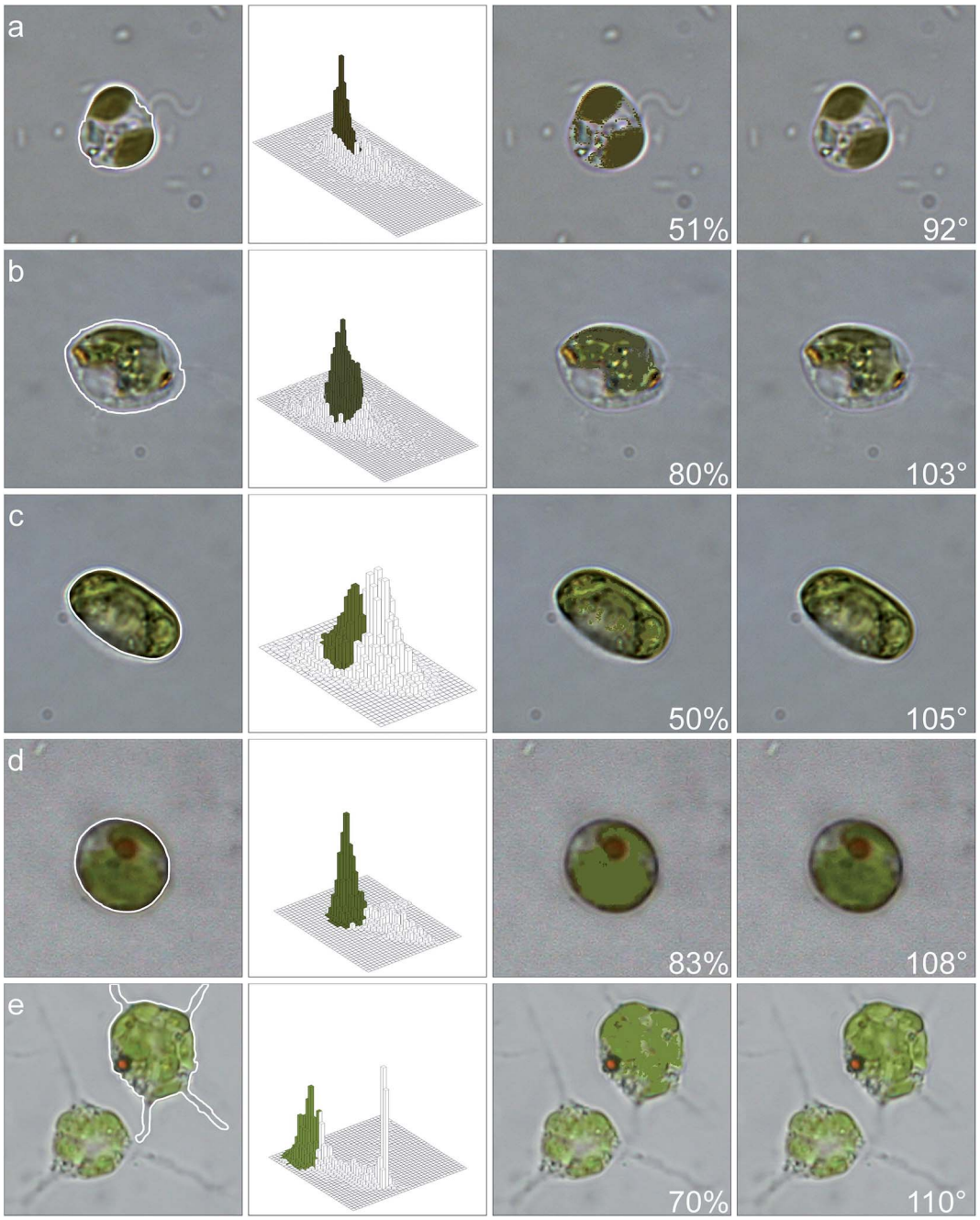


**Fig. 10** List of microalgae ordered by characteristic color hue (CCH). Range from 92° to 110°.

of the images to the different database classes were verified by a phycology expert. Each database class corresponds to an alga taxonomically recognized by the expert at the genus and/or species level.

Four combinations of features patterns were used:

(1) Centroid distance spectrum and absorption spectrum

(2) Centroid distance spectrum and characteristic color

(3) Dissimilarity measure and absorption spectrum

(4) Dissimilarity measure and characteristic color

These four combinations were processed according to the minimum distance criterion to achieve algae categorization. They produced accuracy values with no significant differences. The best accuracy (97.2%) was reached by the first combination, while the other three combinations gave, respectively 96.6%, 95.8% and 94.7% accuracy (Fig. 12).

All these accuracy values are higher than those obtained by other identification and classification methods cited in the Introduction, indicating on the one hand that the choice of the new features we adopted is effective for classifying algae, and on the other hand that color information (*i.e.* pigment signature) is essential to achieve reliable classification results. Therefore, we can state that the characteristic colour calculated in our

 

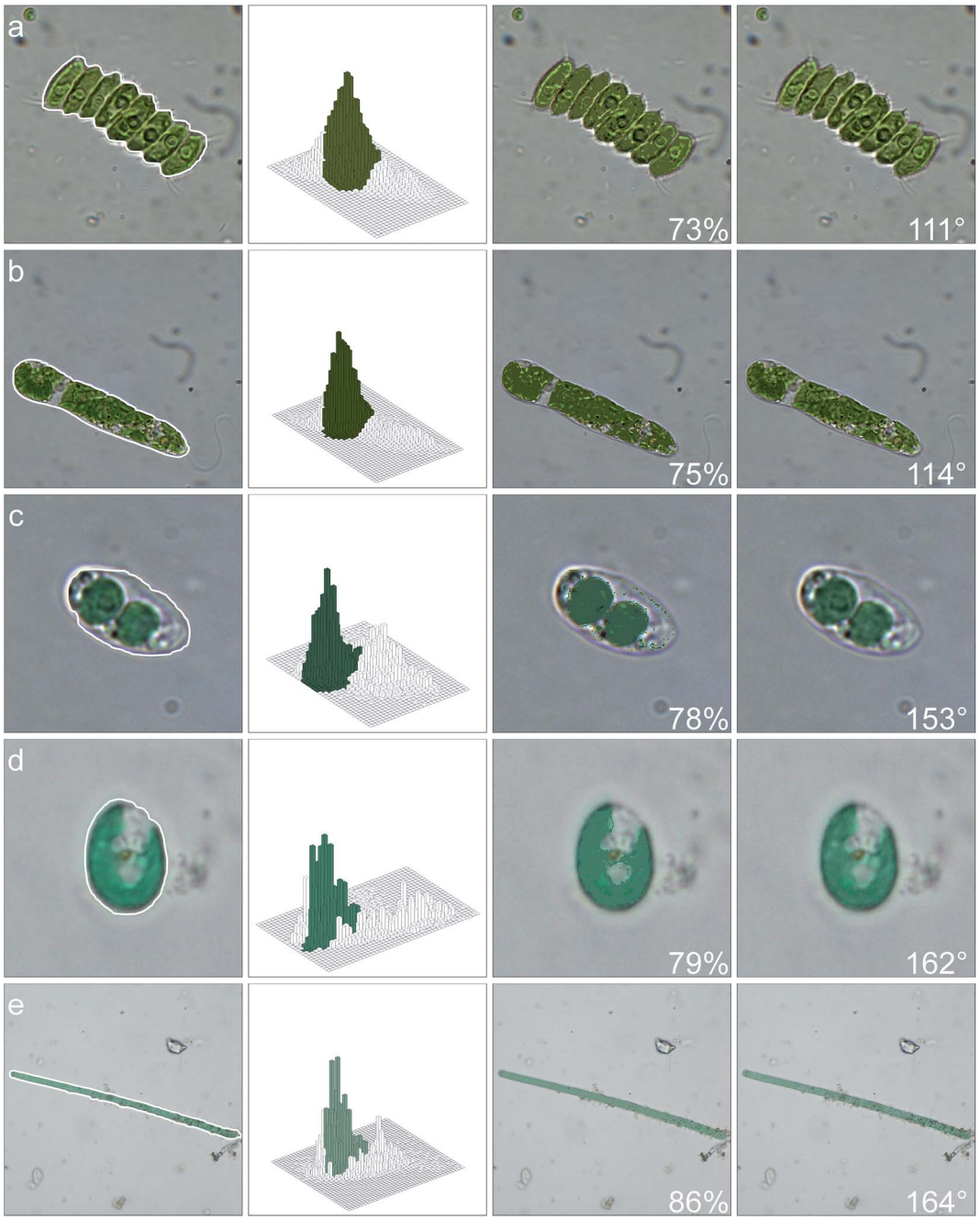


**Fig. 11** List of microalgae ordered by characteristic color hue (CCH). Range from 111° to 164°.

procedure contains visual information equivalent to those that can be obtained by absorption spectroscopy, hence it really represents the pigment fingerprint of the alga. Absorption microspectroscopy is perfectly suited to gather colour information from algal samples, but it is not a routinely adopted methodology. Microspectroscopy set-ups[34] need to be coupled to an image processing set-up to integrate colour information with morphological information in order to obtain usable data. With this configuration, data processing is clearly off-line, much more expensive, and time-consuming.

The novelty represented by our methodology is the capacity to obtain real time accuracy values comparable to those obtained by a phycologist, introducing a categorization parameter that has never been used before, *i.e.* the pigment fingerprint of each alga, combined with a very simple technology.

Though, our methodology at the present stage of development cannot be considered ready for field analysis applications, mainly because of the lack of hardware automations, it is still very promising for future automated systems.

| Features Pattern | Length of Features Pattern | Classification Accuracy (%) |
|---|---|---|
| Centroid Distance Spectrum + Absorbance Spectrum | 360 + 601 | 97.2 |
| Centroid Distance Spectrum + Characteristic Color | 360 + 3 | 96.6 |
| Dissimilarity Measure + Absorbance Spectrum | 1 + 601 | 95.8 |
| Dissimilarity Measure + Characteristic Color | 1 + 3 | 94.7 |

**Fig. 12** Classification accuracy with different features patterns.

## References


1 L. Barsanti and P. Gualtieri, *Algae: anatomy, biochemistry, and biotechnology*, CRC Press, Boca Raton, 2006.
2 M. D. Guiry, *J. Phycol.*, 2012, **48**, 1057–1063.
3 D. M. Kocak, N. D. Lobo and E. Widder, *IEEE J. Oceanic Eng.*, 1999, **24**, 81–95.
4 G. J. Kirkpatrick, D. F. Millie, M. A. Moline and O. Schofield, *Limnol. Oceanogr.*, 2000, **45**, 467–471.
5 J. Seppälä and M. Balode, *Hydrobiologia*, 1997, **363**, 207–217.
6 C. S. Yentsch and D. A. Phinney, *J. Plankton Res.*, 1985, **7**, 617–632.
7 K. Furuya, M. Hayashi, Y. Yabushita and A. Ishikawa, *Deep-Sea Res., Part II*, 2003, **50**, 367–387.
8 M. D. Mackey, D. J. Mackey, H. W. Higgins and S. W. Wright, *Mar. Ecol.: Prog. Ser.*, 1996, **144**, 265–283.
9 L. H. Zhang, J. Zhang and M. Cao, *Mar. Sci.*, 2002, **26**, 60–65.
10 C. A. Scholin, R. Marin, P. E. Miller, G. J. Doucette, C. L. Powell, P. Haydock, J. Howard and J. Ray, *J. Phycol.*, 1999, **35**, 1356–1367.
11 G. Gorsky, P. Guilbert and E. Valenta, *Mar. Ecol.: Prog. Ser.*, 1989, **58**, 133–142.
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 Comput. Appl.*, 1997, **5**, 99–105.
14 K. V. Embleton, C. E. Gibson and S. I. Heaney, *J. Plankton Res.*, 2003, **25**, 669–681.
15 M. B. Blaschko, G. Holness, M. A. Mattar, D. Lisin, P. E. Utgoff, A. R. Hanson, H. Schultz, E. M. Riseman, M. E. Sieracki, W. M. Balch and B. Tupper, *7th IEEE Workshop on Applications of Computer Vision (WACV 2005)*, Breckenridge, CO, USA, 2005.
16 H. Mansoor, M. Sorayya, S. Aishah and A. A. M. Mogeeb, *International Conference on Environmental and Computer Science*, Singapore, 2011.
17 N. Blackburn, A. Hagström, J. Wikner, R. Cuadros-Hansson and P. K. Bjørnsen, *Appl. Environ. Microbiol.*, 1998, **64**, 3246–3255.
18 M. F. Wilkins, L. Boddy, C. W. Morris and R. R. Jonker, *Appl. Environ. Microbiol.*, 1999, **65**, 4404–4410.
19 S. B. Kamath, S. Chidambar, B. R. Brinda, M. A. Kumar, R. Sarada and G. A. Ravishankar, *Biosens. Bioelectron.*, 2005, **21**, 768–773.
20 J. Cheng, G. Ji, C. Feng and H. Zheng, *International Conference on Information Technology and Computer Science*, Kiev, Ukraine, 2009.
21 Z. Yao, M. Fei, K. Li, H. Kong and B. Zhao, *Neurocomputing*, 2007, **70**, 641–647.
22 L. Zhang, Z. Luo, B. Wang and J. Zhang, *International Conference on Electric Informationand Control Engineering (ICEICE)*, 2011.
23 K. Rodenacker, B. Hense, U. Jütting and P. Gais, *Microsc. Res. Tech.*, 2006, **69**, 708–720.
24 R. Thar, M. Kühl and G. Holst, *Appl. Environ. Microbiol.*, 2001, **67**, 2823–2828.
25 H. Xupeng, S. Rongguo, F. Zhang, X. Wang, H. Wang and Z. Zheng, *J. Ocean Univ. China*, 2010, **9**, 16–24.
26 E. Trampe, J. Kolbowski, U. Schreiber and M. Kühl, *Mar. Biol.*, 2011, **158**, 1667–1675.
27 H. M. Sosik and R. J. Olson, *Limnol. Oceanogr.: Methods*, 2007, **5**, 204–216.
28 S. Cuiping, Y. Chenhui, L. Huizhen and K. Lin, *Second International Conference on IEEE*, 2010.
29 H. du Buf and M. M. Bayer, in *Automatic Diatom Identification*, ed. H. d. Buf and M. M. Bayer, World Scientific Publishing Company, Singapore, Editon edn, 2002, vol. 51, pp. 289–298.
30 A. Verikas, A. Gelzinis, M. Bacauskiene, I. Olenina, S. Olenin and E. Vaiciukynas, *Expert Syst. Appl.*, 2012, **39**, 6069–6077.
31 M. Mosleh, H. Manssor, S. Malek, P. Milow and A. Salleh, *BMC Bioinf.*, 2012, **13**, S25.
32 V. Evangelista, M. Evangelisti, L. Barsanti, A. M. Frassanito, V. Passarelli and P. Gualtieri, *Int. J. Biol. Sci.*, 2007, **3**, 251–256.
33 R. C. Gonzalez, R. E. Woods and S. L. Eddins, *Digital Image Processing Using MATLAB*, Pearson/Prentice Hall, London, 2004.
34 L. Barsanti, V. Evangelista, A. M. Frassanito, N. Vesentini, V. Passarelli and P. Gualtieri, *Micron*, 2007, **38**, 197–213.
35 H. W. Zieler, *The optical performance of the light microscope part 1*, Microsope Publications Ltd., London, 1972.
36 P. Coltelli and P. Gualtieri, *Int. J. Bio-Med. Comput.*, 1990, **25**, 169–176.
37 A. Ford and A. Roberts, *Colour space conversions*, Westminster University, London, 1982.


 