Fluorescence-assisted image analysis of freshwater microalgae

Ross F. Walker *, Kanako Ishikawa, Michio Kumagai

Lake Biwa Research Institute, 1-10 Uchidehama, Otsu, Shiga 520-0806, Japan

Abstract

We exploit a property of microalgae—that of their ability to autofluoresce when exposed to epifluorescence illumination—to tackle the problem of detecting and analysing microalgae in sediment samples containing complex scenes. We have added fluorescence excitation to the hardware portion of our microalgae image processing system. We quantitatively measured 120 characteristics of each object detected through fluorescence excitation, and used an optimized subset of these characteristics for later automated analysis and species classification. All specimens used for training and testing our system came from natural populations found in Lake Biwa, Japan. Without the use of fluorescence excitation, automated analysis of images containing algae specimens in sediment is near impossible. We also used fluorescence imaging to target microalgae in water samples containing large numbers of obtrusive nontargeted objects, which would otherwise slow processing speed and decrease species analysis and classification accuracy. Object drift problems associated with the necessity to use both a fluorescence and greyscale image of each microscope scene were solved using techniques such as template matching and a novel form of automated seeded region growing (SRG). Our system proved to be not only user-friendly, but also highly accurate in classifying two major genera of microalgae found in Lake Biwa—the cyanobacteria *Anabaena* spp. and *Microcystis* spp. Classification accuracy was measured to be over 97%.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Image processing; Microalgae; Phytoplankton; Cyanobacteria; Classification; Sediment analysis; Fluorescence

1. Introduction

The analysis of microalgae in sediment samples presents the phycologist with several challenges, one being the difficulty in locating specimens among the very 'busy' microscope scenes. For the same reason, implementing an automated image processing system as a tool for sediment analysis presents the image analyst with a similar problem—how can an image of such a complex sediment scene (see Fig. 1a) be successfully segmented (separated into regions) and labelled as "microalgae" or "other" parts?

While image analysis has been used as a tool for microalgae analysis for over two decades, its application to detecting and analysing microalgae in sediment samples had yet to be conducted. This is because the complex scenes presented by sediment samples cannot be easily analysed using high-resolution greyscale imaging and conventional light microscopy. The main

---

* Corresponding author. Tel.: +81-77-526-4690; fax: +81-77-526-4803.
E-mail addresses: walker@lbri.go.jp (R.F. Walker), ishikawa@lbri.go.jp (K. Ishikawa), kumagai@lbri.go.jp (M. Kumagai).
URL: http://www.members.tripod.com/~Dr_Ross_F_Walker.

difficulty is the inability to locate and accurately segment microalgae image objects while they are surrounded (or even occluded) by sediment and detritus—a result of the microalgae and sediment having similar greyscale intensities.

A solution to this problem can be found in a characteristic possessed by microalgae—that of their ability to fluoresce when exposed to light of specific frequencies (Glazer, 1987; Mur et al., 1999; Whitton, 2000; Herman, 1998; Sharma and Schulman, 1999). The following work details an automated image analysis system that exploits this property. We have developed a fluorescence-assisted image processing and classification system for analysing microalgae specimens in "busy" images. By "busy", we mean images that contain many nontargeted objects (dirt, zooplankton, other water-born debris, etc.) that are normally difficult or impossible to process using conventional greyscale image processing. Because of the autofluorescence properties of microalgae, it is possible to automatically locate and analyse them without processing all other objects in the image.

As can be seen in Fig. 1, using fluorescence can be a very successful technique for analysing microalgae colonies and other autofluorescing structures hidden in sediment samples. Fluorescence excitation clearly shows the locations of any microalgae objects within the sediment, allowing accurate location and subsequent segmentation (separation of the object from its surrounding background). We have implemented fluorescence excitation in the hardware portion of our phycological image processing system (see Walker and Kumagai, 2000). Without the use of fluorescence excitation, automated analysis of phycological images would be near impossible.

As well as sediment analysis, we have also used fluorescence excitation and image processing to aid in the automated detection, analysis, and species classification of microalgae in water samples. Such samples often contain a large diversity of water-born organisms, such as detritus, sediment, zooplankton, as well as microalgae. When targeting a specific microalgae species for analysis using conventional bright-field microscopy, these other nontargeted objects can both slow down analysis and increase analysis errors. This is because the analysis system does not know a priori which objects are microalgae, thus all objects must be processed. By using fluorescence to assist our analysis

and classification of microalgae, we can reduce the number of image objects that need to be processed, simply by targeting only the objects that fluoresce. Moreover, for images that contain no microalgae objects, no processing is required because no autofluorescence occurs. This can result in a considerable reduction in processed data, effectively improving the *speed* of our analysis system. Furthermore, the number of species requiring classification is now significantly reduced, allowing a simplified classifier design and corresponding decrease in system classification *error rate*.

2. Problems with fluorescence-assisted image analysis

Using fluorescence to aid the automated analysis of microalgae presents the researcher with several challenges, mainly due to technological limitations in currently available hardware:

1. A fluorescence image (an image of fluorescing objects, as shown in Fig. 1b) often contains little useable analytic information for species identification, necessitating also the use of a greyscale image;
2. A greyscale image (an image of nonfluorescing objects) captured by a fluorescence camera has low spatial resolution;
3. Both fluorescence and greyscale images cannot be captured simultaneously.

(1) Firstly, the fluorescence image itself contains little useable analytical information for species identification. Fluorescence images tend to have low photometric resolution (low contrast) as well as low spatial resolution caused by limitations of the fluorescence camera (explained below). Adding to the problem is the fact that a fluorescing object acts as a mass of omnidirectional point sources of light, resulting in a blurring effect when imaged by a microscope, which generally requires parallel light for optimal image resolution. To overcome this limitation, a greyscale image of the same scene is necessary if accurate object classification is required. The fluorescence image is used to locate only fluorescing objects (microalgae) in the greyscale image, by segmenting

(a) **(b)**
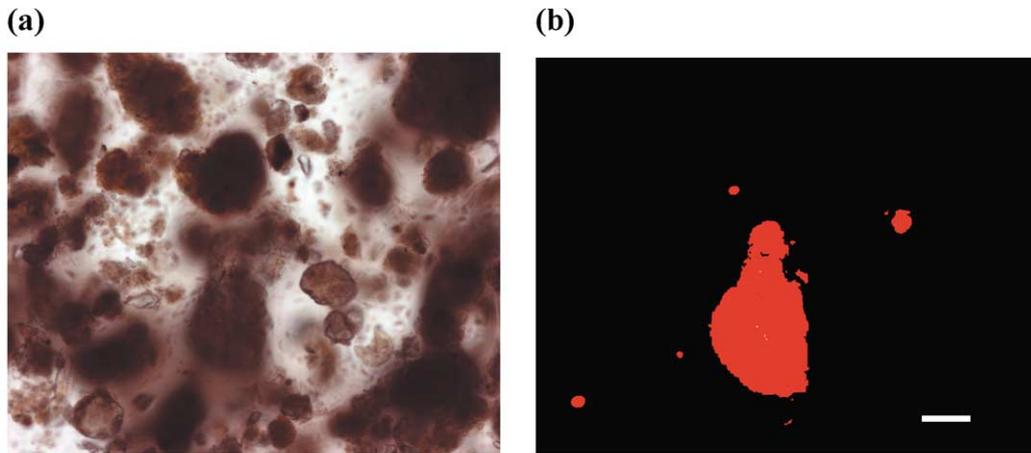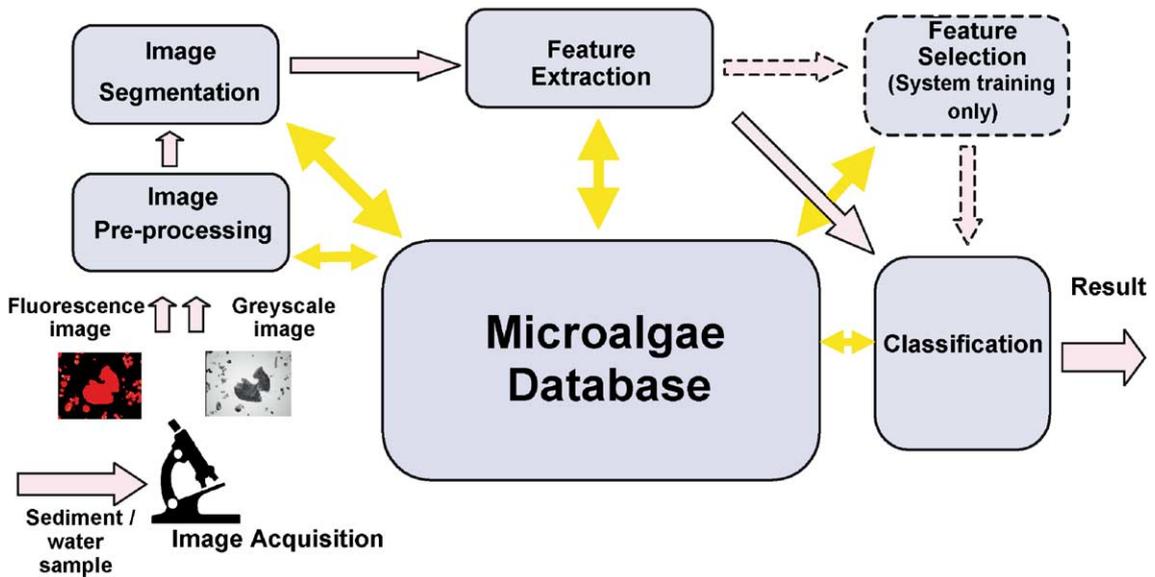
Fig. 1.

Fig. 3.

Fig. 1. Image of Lake Biwa sediment, containing almost completely obscured microalgae specimens (a); corresponding fluorescence image (b). Scale bar = 100 μm.

Fig. 3. Image processing system block for fluorescence-assisted microalgae detection and analysis. Note that feature selection is only undertaken during the training phase of the sytem.

those areas of the greyscale image that have corresponding fluorescing parts.

(2) Unfortunately, using a fluorescence camera to take a greyscale image does not completely solve the spatial resolution problem—see Fig. 2. A fluorescence camera is optimised for high photometric sensitivity to facilitate low-light imaging, and for this reason, spatial resolution suffers. Using a greyscale image captured by a fluorescence camera allows classification to about the genus level—species-level classification accuracy for visually similar species such as *Microcystis aeruginosa* and *M. wesenbergii* would be quite poor. If *accurate*
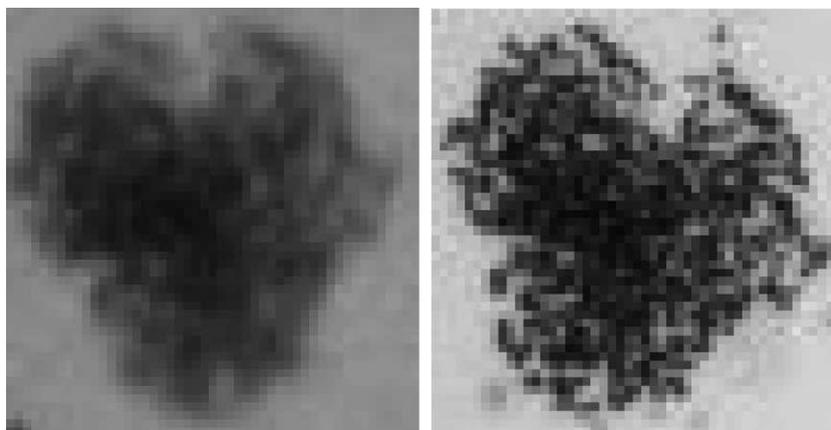
Fig. 2. Magnified view of greyscale image tile of a small *Microcystis* specimen captured by a fluorescence camera (left), and the same specimen captured by a high-resolution greyscale camera (right). Notice the greater spatial resolution of the greyscale camera, allowing individual colony cells to be discerned. The lower resolution fluorescence image displays a typical 'blurred' or 'smoothed' photometric response. Note: both greyscale scenes shown in this figure are small 85×85 μm sub-images of the microscope's much larger 2.2×1.8 mm field of view.

species-level classification is required, it is thus necessary to capture both fluorescence *and* high-resolution greyscale image of each scene using separate cameras. Such a dual-camera system is currently unavailable, so we will limit our discussion to the case of using fluorescence and greyscale images captured by a single fluorescence camera for genus-level classification.

(3) Simultaneous capture of both fluorescence and greyscale images is difficult, if not impossible, using current microscope technology, especially when using a single fluorescence camera. A delay of several seconds occurs between the capture of the two images. During this time, specimens within a water sample can move, resulting in a registration mismatch between the fluorescing object and its corresponding greyscale counterpart (see Fig. 4). The type and amount of movement is different for each image object, making a solution to this registration problem difficult.

Autofluorescence signals from nonmicroalgal sources such as detritus and litter cannot be avoided and as a result can increase the complexity of the scenes to be processed. However, such objects generally do not pose a significant detriment to system performance because they can be detected as non-algal bodies via their morphometric and textural properties during the pattern recognition and classification process.

In the subsequent section, we will introduce the hardware and software technology used in our fluorescence-assisted image analysis system, and present a

solution methodology for the problem of multi-component registration mismatch discussed previously. Finally, we will evaluate our system's performance when applied to two types of analysis:

1. Analysis of sediment samples—locating, segmenting, and measuring statistical properties of microalgae contained in sediment samples from Lake Biwa;
2. Analysis of water samples—locating, segmenting, measuring statistical properties, and genus-level classification of microalgae suspended in water samples from Lake Biwa.

3. Methods

3.1. Specimen collection and treatment

Samples were obtained from August to September 1999 from the Kitayamada fishery port in the South Basin of Lake Biwa (35°1′ 26″ N, 135°55′ 5″ E)—a shallow eutrophic area of the lake (ca. 2.5 m depth). Toxic cyanobacterial blooms regularly occur there in the summer months (Shiga Prefectural Government, 2000). Sediment core samples were taken using an undisturbed core sampler (HR Rigo) and stored in a refrigerator (4°). Subsamples (1 ml) from the top surface layer (0–2 cm) were diluted with 50 ml water.

Water samples were taken from the lake water surface using a sampling bucket and immediately brought back to the laboratory. Sediment solution and water samples (1 ml) were poured onto an Utermöhl sedimentation chamber to take pairs of fluorescence and greyscale digital images.

3.2. System hardware

The complete system can be broken down into two main functional components:

- Epifluorescence and bright-field imaging system for sediment and water analysis, and
- Image processing and database system.

The imaging system consists of a digital colour CCD camera for fluorescence use (Hamamatsu C4742-95-12SC) mounted on an inverted epifluorescence microscope (Olympus IX70) fitted with filters for Chlorophyll fluorescence observation—Olympus WIG cube (exciter filter [band-pass]: BP520–550, dichroic mirror: DM565, barrier filter [low-pass]: BA580IF). Both fluorescence *and* greyscale images were taken of each sediment or water scene by TWAIN-compatible capture software running on a high-speed personal computer. Images had a spatial domain of 1280×1024 pixels at a spatial resolution of approximately 1.7 μm/pixel when using a $4 \times$ microscope objective lens. Each fluorescence/greyscale image pair was stored on hard disk in a microalgae database (see Fig. 3). Statistical characteristics (morphometric, texture, frequency domain parameters) of each microalgae object were quantitatively measured and used to determine taxonometric classification. To provide a robust test of our system, we used species of cyanobacteria, which are readily available in the eutrophic areas of Lake Biwa's south basin. Sediment and water samples were collected and processed during the summer bloom period of August to October 1999. The fluorescence image database contained over 600 microalgae objects whose taxonomy has been individually determined by a phycologist. The database currently comprises predominantly of five species of cyanobacteria—*M. aeruginosa*, *M. wesenbergii*, *Anabaena smithii*, *A. ucrainica*, and *Aphanizomenon flos-aquae*. The analysis and classification of microalgae samples is totally automated except for fluorescence/greyscale switching of the microscope and camera, which is done manually until automated switching technology becomes available.

4. Image processing methodology

Fig. 3 details the architectural layout of the image processing system used for microalgae detection and analysis. In the following sections, we will discuss the functionality of each processing block, especially those related to fluorescence imaging. We refer the reader to Walker and Kumagai (2000) for a more detailed discussion of our image processing techniques.

4.1. Image preprocessing—nonuniform illumination correction

Digital greyscale images received for analysis are preprocessed to reduce the effects of nonuniform illumination. Such an undesirable characteristic, which represents a nonlinear transformation of the true image intensity data, can present a severe challenge to subsequent image processing algorithms such as segmentation. In the present system, the microscope's illumination transfer function is measured and used to correct the greyscale images. Illumination correction is not required for the fluorescence images.

4.2. Image segmentation

Objects within each image are separated from the image background via the process called *segmentation*. By *object*, we mean any body (group of image pixels) that appears darker than the image background. This usually can include dirt, non-alga species, detritus, etc., as well as microalgae. However, using fluorescence allows the system to avoid segmenting such non-algae objects. We achieved segmentation by processing both a fluorescence and a greyscale image of the same scene. A binary segmentation 'mask' is first formed from the fluorescence image. This mask contains transparent and nontransparent areas, corresponding to phycological and nonphycological objects, respectively. The mask is then overlayed on the original greyscale image, effectively using it as a 'window' into the greyscale image. Areas of the greyscale image that show through the mask

(the phycological objects) are then removed from the greyscale image and processed. Because both the fluorescence and greyscale images are of the same microscope field, there should be a one-to-one correspondence between each of the mask's 'windows' and the microalgae objects in the greyscale image.

Initially, a rough mask is formed by a simple thresholding of the fluorescence image at an intensity level automatically determined by its intensity statistics. This mask is then further processed to remove small imaging 'noise' particles or other objects smaller in size than the minimum expected size of the targeted microalgae species. Finally, mask edge pixels are smoothed to form more uniform object boundaries. These steps are performed via a series of morphological image processing algorithms based on mathematical morphology (Serra, 1982; Vincent and Beucher, 1989).

4.3. Image registration solution methodology for water sample analysis

Phycological objects that are suspended in water can change their spatial position and orientation over time due to some inbuilt ability, gravity, or water currents formed via heating of the water sample by the microscope's light source. For similar reasons, it is also possible for the shape of objects to change (deform). This is particularly true when water currents act on fragile microalgae such as the *Anabaena* genus. As mentioned previously, these temporal changes occurring in the microscope's field-of-view can have

a detrimental impact on the segmentation accuracy of the system, because it is impossible to simultaneously capture both fluorescence and greyscale images of the same microscope scene, using currently available microscope and imaging technology.

Although such spatial translation and deformation are generally of limited magnitude over the time required to capture both fluorescence and greyscale images, it is nevertheless necessary to correct such changes if accurate segmentation of image objects is necessary.

Fig. 4 shows two examples of gross segmentation mask misregistration—a result of such microalgae movement during the delay between capturing the fluorescence and greyscale images. This movement can be modelled as a process involving the following three local image spatial transformations:

- Translation—movement of the object in the spatial (x and/or y axis) domain;
- Rotation—rotation of the object around its centre of gravity;
- Deformation—changes to the physical shape of the object.

We investigated possible solutions to the problem of fluorescence/greyscale image misregistration caused by these multi-component spatial transformations.

One method of correcting spatial translation is by a process called 'template matching' (Gonzalez and Woods, 1993) using the 'cross-correlation' function (Castleman, 1996). Template matching, however, is
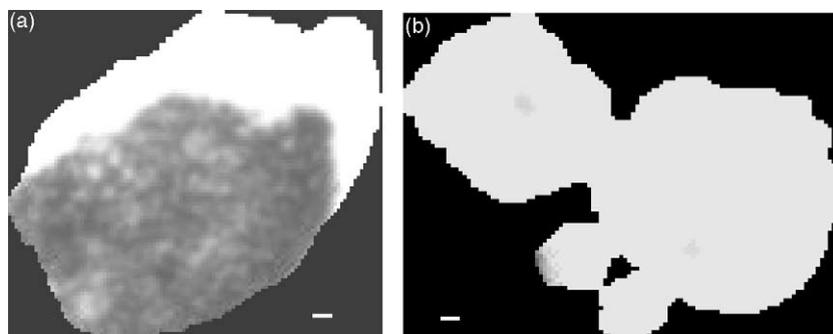
Fig. 4. Image registration problem as a result of specimen movement. Here, we show two examples of the binary fluorescence mask image being overlayed on the greyscale image at the original location of the specimen. Because of specimen movement, partial segmentation mismatch (a) or even complete segmentation mismatch (b) can occur. Note: both fluorescence mask and greyscale scenes shown in this figure are sub-images of the microscope's much larger field of view. Scale bar = 20 μm.

most effective for pure translations—it cannot solve the problems of rotation and deformation, thus any rotational or deformational effects will remain. It is possible to combine an iterative process of both cross-correlation and fluorescence mask rotation to solve the problem of translation and rotation. However, this approach is far too slow (i.e. computationally taxing) due to the large number of possible rotations involved. Also, this combination cannot provide a solution to object deformation.

We felt that an alternative technique called *Region-based segmentation*, in combination with template matching, could provide an adequate solution to all three spatial transformations. Region-based segmentation techniques (Young, 1994), including 'region growing', assume image components are spatially coherent or homogeneous, and separated by sharp boundary discontinuities. For microalgae species, changes in shape (deformation) alone will not result in large spatial coherence changes (changes to the internal structure of the specimen). It should therefore be possible to use correlation techniques to determine the translational component of an object's spatial movement, and then use (say) *seeded region growing* (Adams and Bischof, 1994) to solve the problem of rotation and deformation. The centre of the template-matched fluorescence object mask can be used as a 'seed' to mark a point internal to the greyscale object to be segmented. Region growing would then be used to grow this seed within the greyscale image object, based on homogeneity criteria. Following successful segmentation, characteristics such as object size, shape, texture, etc., can be automatically measured and statistically analysed. Using statistical pattern analysis (Fukunaga, 1990; Hand, 1981), the genus or species membership of the object is determined. The resulting system would then provide fully automated microalgae species classification of highly complex water and sediment samples. We implemented both template matching and region growing techniques within our segmentation algorithm. In the following paragraphs, we describe the algorithm implementation and the result of their use in our fluorescence-excitation microalgae image analyser.

4.3.1. Template matching

For each object in the fluorescence binary image mask, we extract a small image tile just large enough

to contain the entire spatial domain of the object. We then template-match this tile to the greyscale image by *sliding* the tile over the greyscale image, and applying the cross-correlation operator:

$$C(r, c) = \sum_{x,y} I(r + x, c + y)S(x, y) \qquad (1)$$

where I represents the greyscale image, S represents the fluorescence image segmentation mask tile of the object to be segmented, and $C(r,c)$ is the cross-correlation product at spatial displacement (r,c). At each tile position, a local cross-correlation between the segmentation tile and greyscale image neighbourhood was performed. The cross-correlation value at image tile displacement r,c (Eq. (1)) equals the summation of the point-by-point product of the binary fluorescence mask tile and greyscale image intensities across the domain of the tile. Because an image object is darker than the surrounding background, the tile position with the *minimum* cross-correlation product represents the location of the translated object in the greyscale image.

To increase the speed of the algorithm, the domain of the slide area was limited to an area 50 pixels around the original tile size. For example, a 100×100 pixel segmentation image tile extracted from the fluorescence image would be slid over a 200×200 pixel local neighbourhood within the greyscale image. The size of this 'search' neighbourhood was determined by observing the maximum spatial movement of objects when processing water samples—movement larger than 50 pixels was not observed.

4.3.2. Region growing

Once an approximate location of the object in the greyscale image is determined, seeded region growing (SRG) is commenced. In the following discussion, we use the terms 'point' and 'pixel' interchangeably. First, a *seed point* within the object is determined, from which the region will be 'grown'. To ensure that the seed point is located within the object, we select the seed point as being the pixel with the lowest greyscale intensity located within the bounds of the template-matched segmentation mask. Because an image object is darker than the surrounding background, choosing the lowest pixel intensity guarantees a seed point within the object. Neighbouring pixels of

this seed point that are 'similar' in characteristic (explained later) are then added to a pixel queue.

After selecting this start point, an iterative process commences:

- Remove point from front of pixel queue and update *region image*;
- Locate the eight nearest neighbours of this point;
- Add to the queue those neighbouring points that are not already in the queue, have not been processed yet, and which have sufficient *similarity*;
- Sort the queue points in order of similarity;
- Repeat until queue is empty.

The 'region image' I_R is simply a new greyscale image that will eventually contain the segmented phycological object. This image initially contains only one point (the seed point), and gradually grows until the iterative process is finished (the queue is empty). *Similarity* is a measure of how similar a point's intensity is to the current *average* intensity of the 'region image',

$$S(r,c) = \text{abs}(I(r,c) - \bar{I}_R),$$

where S is the similarity measure for pixel (r,c), $I(r,c)$ is the pixel intensity at point (r,c), and $\bar{I}_R$ is the average intensity of the region image. As points are added to the region image, the intensity average $\bar{I}_R$ is updated.

New neighbouring points are only added to the queue if their similarity value meets this requirement:

$$S(r,c) \leq (I_{BG} - \bar{I})/2,$$

where I_{BG} is the average intensity of the image background. That is, if the pixel is closer in intensity to the

object than the image background, it is considered a part of the object and added to the pixel queue for later processing.

Fig. 5 details the growth of a region using SRG. The targeted object to be extracted is a *Microcystis* specimen. The image sequence shows the gradual growth of the region from a single pixel to the entire specimen image. Note in the final image how the algorithm has, at two locations, correctly avoided the brighter background pixels contained within the specimen. The inclusion of such false object points leads to inaccurate measurement of the object's statistical properties such as area, and ultimately can result in incorrect species classification of the specimen.

It should be noted that our region growing procedure is a modified form of that of Adams and Bischof (1994). The latter algorithm partitions the entire image into two or more regions, and therefore processes all image pixels. Because the image background is of no interest to us, we choose to only add foreground object pixels to the queue. This modification makes a drastic change to algorithm speed, as the usually large number of background pixels in an image are not processed. Objects are processed at a speed of several hundredths of a second for small objects, to approximately 1 s for the largest objects whose size approached that of the entire microscope field-of-view (such as some *Microcystis* specimens), using an 850-MHz Windows PC.

4.4. Focus check

We individually confirm the focal quality of segmented objects to ensure they possess an adequate level of focus. This step is of vital importance for the case of water sample analysis. Because a slide water sample is a three-dimensional body, some objects may fall outside the in-focus portion of the microscope's
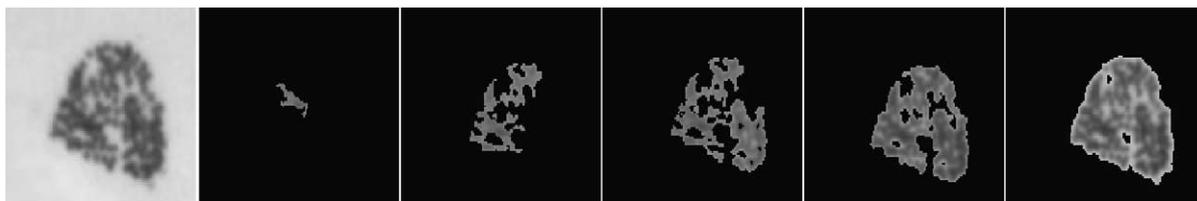
Fig. 5. Seeded region growing in action. This image sequence shows the growth of the region image after 100, 700, 1400, 2100, and 2796 iterations of the region-growing algorithm. The original image of the region to be extracted (a *Microcystis* specimen) is shown on the far left.

view. This is also true for the case of sediment sample images, and it is further compounded by the already poor spatial frequency response of the fluorescence camera. As a result, there can be great variability in the focal accuracy of objects appearing in the microscope's field of view. We fully processed and classified objects with adequate focal quality. Objects that did not achieve a minimum focus limit were analysed to measure simple characteristics such as area and shape, and added to the database, but were not subsequently classified. This is because the defocus effect adversely influences many of the statistical properties of the image, and thus may have a strong negative influence on classification accuracy. We measured focal quality by removing low spatial frequency components of the object image, and averaging the remaining power spectrum across one image dimension (Oliva et al., 1998),

$$F(I) = \sum_r \sum_c (h(r) \otimes I(r,c))^2 \Bigg/ \left(\sum_r \sum_c I(r,c)/A \right)^2,$$

where $F(I)$ is the focal quality measure for greyscale image I of spatial domain $A = r_{max} \times c_{max}$ pixels, $h(r)$ is the spatial domain response of the 1-D high-pass filter kernel, and $\otimes$ is the convolution operator. This is a widely used technique and is computationally light, allowing high-speed focal quality measurement.

4.5. Object feature extraction

When we wish to classify an object into one of several classes, i.e., *Microcystis*, *Anabaena*, etc., we need to quantitatively measure characteristics of the object that may signify its membership class. For instance, the characteristic 'area' is a strong discriminator of class membership when classifying *Microcystis* and *Anabaena* cyanobacteria, as these two genera differ substantially in size. We call these characteristics *features* in pattern recognition terms, and the process of measuring such characteristics as *feature extraction*.

During the training phase of our system, we measured a total of 120 features of each object (see Table 1), including morphometric properties (area,

Table 1
Types of features extracted from each image object

Feature type	Examples	Number measured
Morphometric	Area, circularity	4
Boundary shape	Curvature properties	5
Frequency domain	Fourier components of boundary	14
Second-order statistical properties	Grey Level Co-occurrence Matrix features	97
	Grey Level Variance Matrix features (GLVM—see Yogesan et al., 1994)	

circularity, perimeter length, etc.), object boundary shape properties, frequency domain properties, and second-order spatial statistics including Grey Level Co-occurrence Matrix (GLCM) measures—a powerful method of texture analysis which models texture as an overall or average spatial relationship between grey tones in an image (Haralick et al., 1973; Haralick, 1979; Conners and Harlow, 1980). A complete list of the feature measures used in this research can be found in Walker and Kumagai (2000). For system testing and actual day-to-day operation, only a small optimised subset of these 120 features are measured (see the following section).

4.6. Feature selection

Selecting a subset of discriminatory features from a larger set is called 'feature selection'. The process is arguably one of the most important steps in pattern recognition. Generally, there will exist a high dimensional feature space, with a limited number of data samples to accurately characterise the class distributions within this space. By removing redundant features that do not discriminate between classes, we can better represent this now lower-dimensional space, allowing us to design a more robust classifier. Also, during real-world operation of the system, only this minimal subset of discriminatory features need be measured to classify a specimen.

To find an optimal feature subset, we used a feature selection process called sequential forward-selection/backward-elimination (Hand, 1981). To an initially

empty feature set, our algorithm adds two new features and then removes one feature, repeating iteratively until we have the desired number of optimised features. By doing so, we capture feature pairs that possess higher-order discriminatory power. This method almost always gives optimal results and computationally is comparable to less optimal approaches (Kittler, 1978). Using this method, we found that a total of three features from the original set of 120 possessed sufficient significant discriminatory power to accurately classify the water-born cyanobacteria as *Microcystis* spp. or *Anabaena* spp.

4.7. Classification

Classification was implemented using a general Bayes decision function for assumed Gaussian feature distributions with unequal variance–covariance matrices (Gonzalez and Woods, 1993). The resulting decision surface (where $d_1 = d_2$) is of hyperquadric form:

$$d_i(\mathbf{x}) = \log P_{\omega_i} - \frac{1}{2}\log|C_i| - \frac{1}{2}$$
$$\times \left[(\mathbf{x} - \bar{\mathbf{x}}_i)' C_i^{-1}(\mathbf{x} - \bar{\mathbf{x}}_i)\right], \qquad i = \{1, 2\},$$

where $\mathbf{x}$ is the feature vector of the object to be classified, $d_i(\mathbf{x})$ represents the discriminant measure for $\mathbf{x}$, P_{ω_i} is the a priori probability of class ω_i, and C_i and $\bar{\mathbf{x}}_i$ are the variance–covariance matrix and mean vector, respectively, for class i data (determined from our database of expertly classified microalgae specimens obtained from natural populations).

The resulting decision surface is of hyperquadric. Microalgae in water samples were classified to the genus level. There were insufficient database examples of the species *Aphanizomenon* to provide an accurate statistical model for their feature distributions, so we decided to target only the *Anabaena* and *Microcystis* genera. *A. smithii* and *A. ucrainica* were combined to form the first data class—*Anabaena* spp., while *M. aeruginosa* and *M. wesenbergii* formed the second—*Microcystis* spp.

5. Results and discussion

5.1. Image registration solution for translation, rotation, and deformation

Fig. 6 shows the results of template matching and region growing for an object that had undergone spatial translation and slight deformation after the fluorescence image was captured. The original segmentation mask displacement error is approximately 95% by area (a). This has been mostly corrected by template matching (b). However, some parts of the object have been cropped and many background pixels still exist in the foreground due to deformation of the object. This registration error has been completely corrected by seeded region growing (c).

Fig. 7 is another example showing translation, slight rotation, and deformation changes. The original segmentation mask displacement error is approximately 50% by area (a). This has been corrected by template matching (b); however, slight rotation and
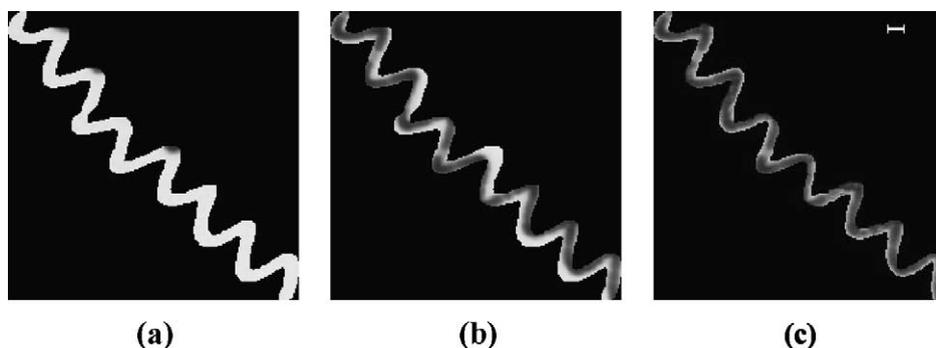
Fig. 6. Original segmented image (a); after template matching (b); after region grown (c). Note the slight shape deformation in image (b), which is corrected in image (c). Scale bar = 20 μm.
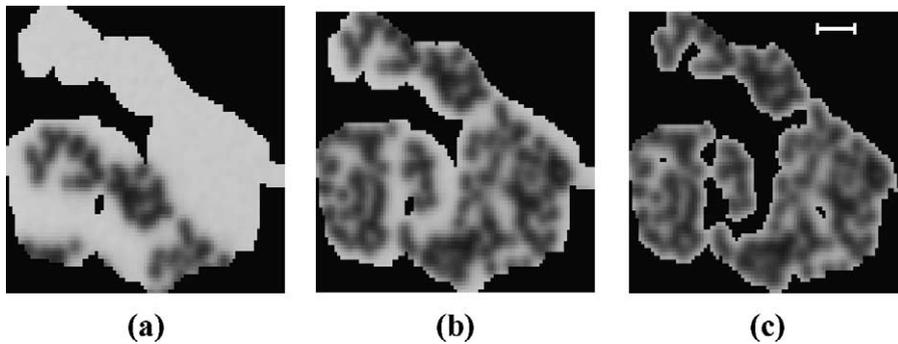
Fig. 7. Original segmented image (a); after template matching (b); after region grown (c). Note the minor shape deformation in image (b), which is corrected in image (c). Scale bar = 20 μm.

deformation error still exists. These errors have been removed by seeded region growing (c).

The success or failure of our image registration correction algorithm is dependent on the amount of spatial rotation and deformation of each object. The method is not affected by pure spatial translation at all (unless the amount of translation is so great that the object falls outside the algorithm's 'search area'—we did not experience any such failures due to large specimen spatial translation). It is possible for the algorithm to fail to successfully locate a specimen due to spatial rotation or deformation, but only if such movement is significant and, even in that case, only if there is another specimen within the search area that better matches the mask of the specimen being template-matched. During system training and evaluation, such template-matching failures were rare to the point of being insignificant.

5.2. Image analysis results

A total of 585 objects were extracted from 112 sediment and water sample images. Our system had no difficulty in locating microalgae specimens located in sediment samples. Measurement of statistical properties such as the number of microalgae objects and object area was extremely rapid, at less than 50 ms per object. Accuracy of measurement depended on two conditions—the strength of the specimen's fluorescence, and whether the specimen was partly occluded by sediment. It should be noted that the accuracy of human screeners would also suffer under the same conditions.

For the water-born microalgae genera *Microcystis* spp. and *Anabaena* spp., an apparent classification rate of approximately 90% was achieved using only one feature (a 'Circularity' measure). Using three features (morphometric and GLCM texture measures), our system was able to classify these images with approximately 97.6% accuracy using the Leave-One-Out classification technique. Fig. 8 shows a plot of apparent classification error vs. the number of features used for classifying the data. Apparent classification error began to increase when more than three optimised features were used. This is most likely a classic case of the 'curse of dimensionality', where the advantage of adding extra features with minimal new discriminatory information is offset by the difficulty in modelling a feature space of greater dimensionality. See Hand (1981, p. 120) and Hand (1997, p. 80) for further discussion of this phenomenon. The confusion matrix in Table 2 shows sufficiently low error rates for both species to provide high confidence in the system's automated analysis capabilities.

5.3. Feature analysis

We analysed the features that were found to provide the highest class-discrimination by the feature selection process. The statistical data for two of these features ('Circularity' and 'Inertia') are shown as a scatter plot in Fig. 9. The third feature is 'Diagonal moment'.

The morphometric feature 'Circularity' is a measure relating object boundary length l to object area A, and is defined as the ratio $l^2/4\pi A$. A perfectly circular
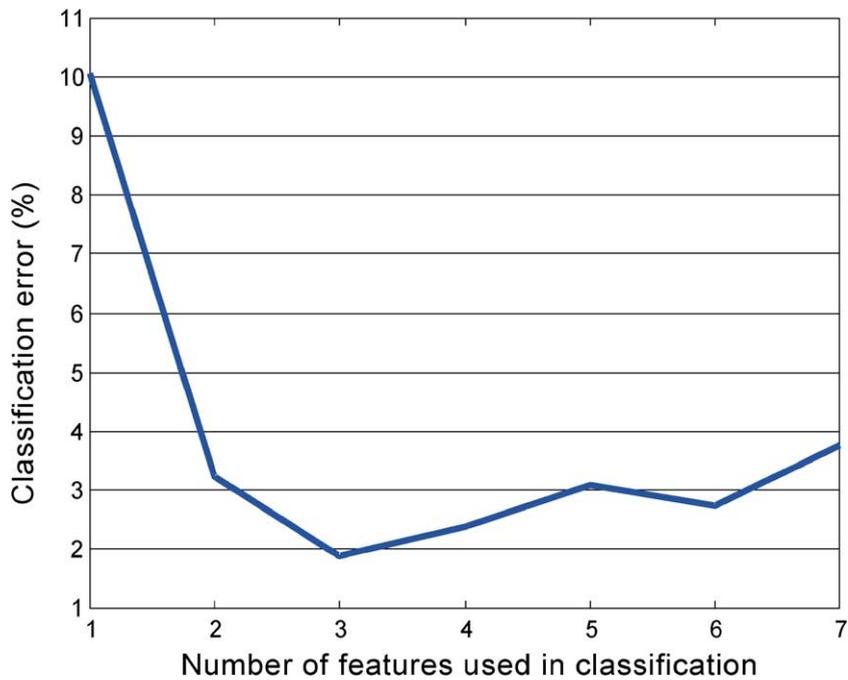
Fig. 8. A plot of system classification error versus the number of statistical features used to discriminate between the *Anabaena* spp. and *Microcystis* spp. genera. We note that using 3 of the 120 measured statistical properties provides the least classification error. Using these 3 features, the system produced an apparent error rate of 1.9% during system design, and a real error rate of 2.73% via Leave-One-Out classification during final performance evaluation.
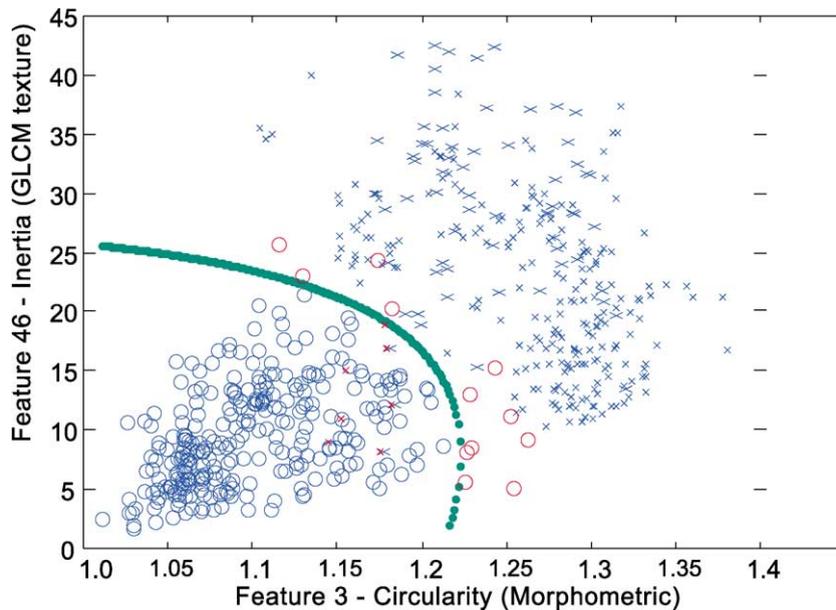
Fig. 9. A scatter plot showing the distribution of object data in the feature space of the best feature pair—'Circularity' and 'Inertia'. The quadratic line that partitions this feature space into the two classes *Anabaena* spp. (×) and *Microcystis* spp. (O) is also drawn.

Table 2
Confusion matrix of classification results for the two data classes of *Anabaena* spp. and *Microcystis* spp.

Actual	Classified as:	
	Anabaena spp. (%)	*Microcystis* spp. (%)
Anabaena spp.	49.1	0.5
Microcystis spp.	1.8	48.5

System performance was evaluated using the Leave-One-Out technique.

object has a circularity measure of 1. *Anabaena* spp. generally has large boundary length nd low internal area, resulting in high circularity measures. *Microcystis* on the other hand are usually more compact in shape, resulting in a lower measures. This relationship can be clearly seen in the scatter plot.

It is often difficult to relate a mathematically derived measurement of an image property to a physical property of the specimen. Such is the case for the GLCM texture feature called 'Inertia'. A full description of this feature and the well-known and often-used GLCM technique in general is beyond the scope of this paper—we refer the reader instead to Haralick (1979) and Conners and Harlow (1980). Qualitatively speaking, the Inertia measure gives high scores to images that contain large variability in intensity over a particular spatial displacement—in this particular case, a displacement of 4 pixels or approximately 7 μm. Images of *Microcystis* specimens generally had large areas with very little variation in intensity, resulting in low Inertia measures. An *Anabaena* specimen on the other hand will generally attain a higher Inertia score, because its long thin shape means that much of its internal structure (dark intensities) is near to its perimeter (light intensities). Once again, this relationship is clearly demonstrated in Fig. 9. The third most discriminatory feature— 'Diagonal Moment'—is another GLCM texture feature. It is very similar to the feature Inertia; however, in this case, it is measuring structure at a finer spatial scale—1 pixel displacement or approximately 2 μm.

6. Conclusions

Fluorescence excitation has been shown to be an invaluable tool for analysing complex image scenes containing microalgae, such as those of sediment samples, which cannot be automatically processed by standard bright-field light microscopy and grey-scale image analysis. Phycological objects in sediment taken from Lake Biwa were easily located and their statistical properties quantitatively measured. The technique also proved invaluable for automated analysis of microalgae contained in water samples. Without fluorescence excitation, an analysis system would need to cope with the vast numbers of non-algae species that co-exist with microalgae, making species classification much more difficult.

Limitations in current microscope technology— that of fluorescence/greyscale imaging switching delay—resulted in our image analysis system being one step away from fully automatic operation, and introduced image segmentation registration error due to specimen movement in water. The three forms of registration error—spatial translation, rotation, and deformation, were corrected using template matching and region growing techniques. We showed that region-growing techniques are invaluable for solving the latter two registration errors. While being computationally intensive, region growing is still less taxing than other techniques based on rotation correction, and simultaneously provides a solution to the difficult problem of segmenting deformed objects. To reduce computational burden, we implemented a computationally light form of seeded region growing which differs significantly from that of Adams and Bischof (1994):

- Only the foreground object is grown;
- The algorithm automatically determines the seed location.

The first difference results in much greater computational speed. The second difference allows the algorithm to be implemented into a fully automated system such as the one being developed at Lake Biwa Research Institute. That is, no human intervention is required to select object seed points—a drawback of the Adams and Bischof method.

Our system is currently unable to cope sufficiently with gross movement of fluorescing specimens such as the energetic Dinophyceae *Ceratium* species, because their locations can change significantly dur-

ing the delay between capturing the fluorescent and greyscale images. However, this deficiency will be overcome once automated switching technology becomes available.

Our system proved to be not only user-friendly, but also highly accurate in classifying two major genera of microalgae found in Lake Biwa—the cyanophytes *Anabaena* spp. and *Microcystis* spp. While only cyanophytes were used in the present study for evaluating system performance, due to their abundance and ease of human expert classification, the system is capable of analysing any type of microalgae at meet the size limitations imposed by the microscope field of view. Classification accuracy was measured to be over 97%. Future improvements in fluorescence imaging technology will allow low-cost, fully automated, species-level analysis and classification of microalgae contained in sediment and water samples.

Acknowledgements

References

Adams, R., Bischof, L., 1994. Seeded region growing. IEEE Trans. Pattern Anal. Mach. Intell. 16 (6), 641–647.

Castleman, K.R., 1996. Digital Image Processing Prentice-Hall, USA.

Conners, R.W., Harlow, C.A., 1980. A theoretical comparison of texture algorithms. IEEE Trans. Pattern Anal. Mach. Intell. PA-MI-2 (3), 204–222.

Fukunaga, K., 1990. Statistical Pattern Recognition. Academic Press, New York.

Glazer, A.N., 1987. Phycobilisomes: assembly and attachment. In: Fay, P., Van Baalen, C. (Eds.), The Cyanobacteria. Elsevier, Oxford, UK, pp. 69–94.

Gonzalez, R.C., Woods, R.E., 1993. Digital Image Processing Addison-Wesley, USA.

Hand, D.J., 1981. Discrimination and Classification. Wiley, USA.

Hand, D.J., 1997. Construction and Assessment of Classification Rules. Wiley, USA.

Haralick, R.M., 1979. Statistical and structural approaches to texture. Proc. IEEE 67 (5), 786–804.

Haralick, R.M., Shanmugan, K., Dinstein, I., 1973. Texture features for image classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621.

Herman, B., 1998. Fluorescence Microscopy. Springer, New York, USA.

Kittler, J., 1978. Feature set search algorithms. In: Chen, C.H. (Ed.), Pattern Recognition and Signal Processing. Sijthoff & Noordhoff, The Netherlands.

Mur, L.R., Skulberg, O.M., Utkilen, H., 1999. Cyanobacteria in the Environment. In: Chorus, I., Bartram, J. (Eds.), Toxic Cyanobacteria in Water. E. & F.N. Spon, London, pp. 15–40.

Oliva, M.A., Bravo-Zanoguera, M., Price, J.H., 1998. Autofocus for phase-contrast microscopy: investigation of causes of non-unimodality. In: Farkas, D.L., Leif, R.C., Tromberg, B.J. (Eds.), Proceedings Optical Investigations of Cells In Vitro and In Vivo vol. 3260. SPIE.

Serra, J., 1982. Image Analysis and Mathematical Morphology. Academic Press, London.

Sharma, A., Schulman, S.G., 1999. Introduction to Fluorescence Spectroscopy. Wiley, New York.

Shiga Prefectural Government, 2000. White Paper on the Environment of Shiga Prefecture (Shiga-ken Kankyo Hakusho Shiryouhen) Otsu, pp. 100–105 (in Japanese).

Vincent, L., Beucher, S., 1989. The morphological approach to segmentation: an introduction. Mathematische Morphologie und Digitale Bildverarbeitung, 25–27 Sep., München.

Walker, R.F., Kumagai, M., 2000. Image analysis as a tool for quantitative phycology: a computational approach to cyanobacterial taxa identification. Limnology 1 (2), 107–115.

Whitton, B.A., 2000. Soils and rice fields. In: Whitton, B.A., Potts, M. (Eds.), The Ecology of Cyanobacteria. Kluwer Academic Publishing, Dordrecht, The Netherlands, pp. 233–255.

Yogesan, K., Albregtsen, F., Danielsen, H.E., 1994. Gray level variance matrix: a new approach to higher order statistical texture analysis. Proc. 3rd Int. Conf. Automation, Robotics, and Comput. Vis., vol. 2, pp. 658–663.

Young, T.Y., 1994. Handbook of Pattern Recognition and Image Processing: Computer Vision, vol. 2. Academic Press, USA.