Procedia Computer Science

Volume 29, 2014, Pages 1981–1992

ICCS 2014. 14th International Conference on Computational Science

Automated Microalgae Image Classification

Sansoen Promdaen[1], Pakaket Wattuya[1], and Nuttha Sanevas[2]

[1] Department of Computer Science, Kasetsart University, Bangkok, Thailand
g5514401596@ku.ac.th, fscipkw@ku.ac.th,
[2] Department of Botany, Kasetsart University, Bangkok, Thailand
fscintsv@ku.ac.th

Abstract

In this paper we present a new method for automated recognition of 12 microalgae that are most commonly found in water resources of Thailand. In order to handle some difficulties encountered in our problem such as unclear algae boundary and noisy background, we proposed a new method for segmenting algae bodies from an image background and proposed a new method for computing texture descriptors from a blurry texture object. Feature combination approach is applied to handle a variation of algae shapes of the same genus. Sequential Minimal Optimization (SMO) is used as a classifier. An experimental result of 97.22% classification accuracy demonstrates an effectiveness of our proposed method.

Keywords: microalgae image classification, microalgae image segmentation, multiple feature combination

1 Introduction

Algae are important microscopic aquatic life forms as they are primary producers in an aquatic food chain and oxygen producers in an aquatic ecosystem. In water resource management, algae are used as a biological index to indicate a quality of water because they are sensitive to environmental changes [17]. Therefore, recognition of microalgae is one of the most important issues in water resource management. However, this task is time-consuming and requires expert biologists to accomplish it. In this work, we proposed a new method for automated classifying and recognizing microalgae in microscopic images.

Despite of its importance, there is a few research works on this problem. For example, the works [11, 21] proposed a classification method for recognizing blue-green algae (also known as cyanobacteria). Blue-green algae are of interest because they present a problem to water quality due to their toxic nature. These methods were proposed to deal with a single division of algae. Our work aims to deal with multiple microalgae divisions (i.e. both harmful and harmless species) that are most commonly found in water resources of Thailand [14, 18]. Twelve genera of microalgae studied in this work are detailed in Table 1. The intended contributions of our work are not limited only to recognize toxic algae for the purpose of water quality assessment, but also to recognize common algae for the purpose of aiding biologists for ecological study of diversity of algae in water resources, and semi-automated

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2014

doi:10.1016/j.procs.2014.05.182

generating microalgae taxonomy. In comparison to the previous works, our work faces with several difficulties as followings:

1. Algae images used in most existing algae recognition system [11, 20, 21] are generated from a similar imaging system, for which a resolution of captured images is known. However, in our work we received microscopic images from various sources, in which different imaging systems and different capturing information are used to produce images. Thus, algae images in our data set have various resolution, illumination and magnification settings. This situation introduces a major problem in a process of algae feature extraction, especially shape measurement features, which will be described later in Section 3.3.

2. In comparison to the previous works [11, 20, 21], in which a single division of blue-green algae has been studied. In our work, we study three divisions of microalgae, namely, blue-green algae, green algae, and euglenoids. Among these divisions, algae in the green algae division are the most difficult to recognize, especially, *Scenedesmus* and *Staurastrum* genera. The algae shapes of each of these genera are much more diverse than the algae shapes of those genera in the blue-green algae division. In this work, we attempt to deal with this problem by combining multiple algae features in a classification process.

3. Some algae in microscopic images do not have clear boundaries. The first cause of the blurred boundary is due to an image acquisition process (e.g. illumination adjustment, wrong focus, etc.). The second cause is from an alga itself. Some algae are enclosed by voluminous gelatinous coat that make true shape boundaries of algae unclear. An example of this situation is illustrated by *Cosmarium* genus as shown in Figure 1. In addition, a transparent appearance of spines and flagellums of algae in a microscopic image makes it difficult to separate them from an image background. The last difficulty of detecting an algae boundary is due to extraneous particles polluted in an image background. When these particles are in contact with algae boundary (as shown in Figure 1, *Phacus* genus), it is mostly impossible to separate them from the algae in a segmentation process. Thus, in this work we propose a new segmentation method that is able to deal with these difficulties. Details of our proposed segmentation method for automatically segmenting algae from an image background will be described in Section 3.2.

4. Some algae in microscopic images do not have clear textures. This problem is critical especially when algae of different genera have similar shape. Hence, the only feature that we can use to indicate the difference between them is their texture. In this work we propose a new method for extracting texture feature from a blurred texture object. The proposed method is based on the texture features proposed by Haralick et al. [7]. Details of the method will be described in Section 3.3.

Studies of microalgae recognition mostly use classification methods, such as SVM [10,20], Artificial Neural Network [11], Support Vector Data Description (SVDD) [20], and Discriminant Analysis [21]. In this work we propose to use the Sequential Minimal Optimization (SMO) algorithm for training a support vector classifier using scaled polynomial kernels [13] in a classification process. SMO has been successfully applied in many application domains, including medical image analysis for lung [4] and brain tumor detection [5], handwritten character recognition, text categorization, and speech recognition [1].

The rest of the paper is organized as follows. In the next section, we described our algae image dataset. A new method for automated algae recognition are described in Section 3 and followed by experimental results in Section 4. Finally, some discussions conclude this paper.

Toxin Type	Division	Genus	Shape
Toxic	Blue-green algae	*Anabaena*	filament
		Oscillatoria	filament
		Microcystis	colony
Non-toxic	Green algae	*Scenedesmus*	colony
		Pediastrum	colony
		Cosmarium	unicells
		Closterium	unicells
		Xanthidium	unicells
		Staurastrum	unicells
		Pleurotaenium	unicells
	Euglenoids	*Euglena*	unicells
		Phacus	unicells

Table 1: Details of twelve genera of microalgae used in our study

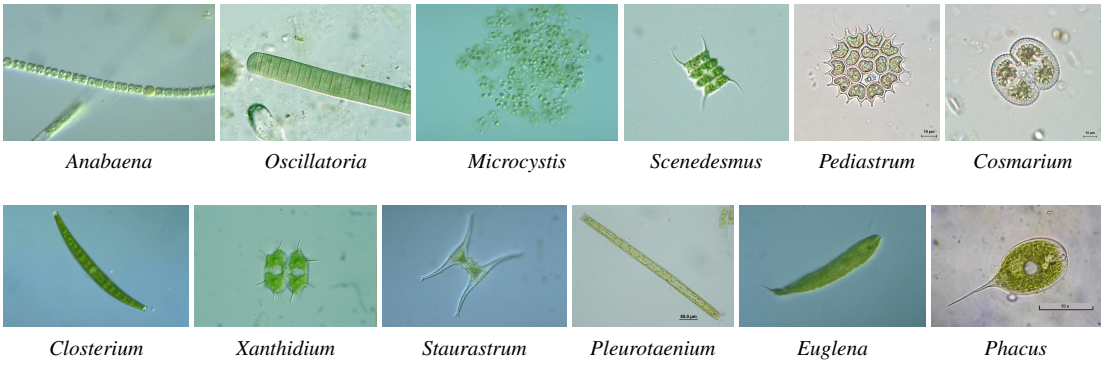


Figure 1: Example of algae images of twelve genera studied in this work

2 Algae Image Dataset

Algae microscopic images in our data set are collected from various sources. The main sources are the research projects [14, 18] conducted by Department of Botany, Kasetsart University. Other sources includes Metropolitan Waterworks Authority, online algae image database and the internet. In this work we study twelve genera of microalgae that most commonly found in water resources of Thailand. The twelve genera are from three divisions, namely, blue-green algae (or cyanobacteria) division, green algae division, and euglenoids division. Details of these genera are summarized in Table 1 and their example algae images are shown in Figure 1. The data set comprises of 720 algae images, 60 images for each genus. Since the images are collected from various sources, their size, illumination setting, and magnification setting are largely varied.

3 Methodology

Classification of algae images consists of 4 main steps, namely, a preprocessing step, image segmentation, feature extraction, and classification.

3.1 Preprocessing

A preprocessing is a process for preparing an input image to be suitable for processing (i.e, segmentation and feature extraction). The first preprocessing process is to resize an input image. Since sizes of our input images are largely varied, we need to resize them into the same scale in order to correctly compute algae shape features, particularly, *shape measurement features* (will be described in Section 3.3). The images whose longest side is larger than 400 pixels are resized to 400 pixels while the other side remains in the same aspect ratio.

The second preprocessing process is to convert a color input image into a gray-scale image. We perform color-to-gray image transformation because we do not use color information of algae in a classification process. The first reason is that our algae images are produced from several imaging systems. Thus, colors of algae of the same genus may be varied significantly, depending on imaging systems and illumination adjustment. Secondly, colors or pigments of algae depend on environmental conditions in which they are growing. Colors of algae of the same genus may vary in a wide range, while colors of algae of different genera may be identical. As a result, color feature of algae is not suitable for identifying or discriminating algae.

3.2 Image Segmentation

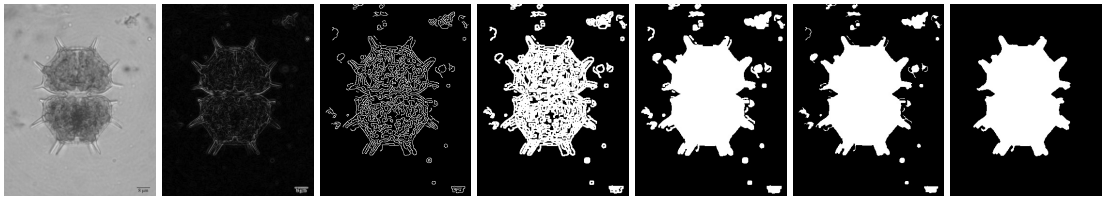
Image segmentation is a process of separating objects of interest from an image background and is of a crucial preprocessing step for most object recognition systems. In general, the accuracy of classification/recognition system depends heavily on the accuracy of object features used in a training process. More precise segmentation result contributes to more accurate object feature computation.

The main difficulties of segmenting algae from an image background are noise and a blurred contour and texture as discussed earlier. Most microscopic images of algae are usually corrupted by noise. Noise in an image can be extraneous materials (or unwanted objects) and illumination artefacts. These noise disrupt a segmentation process and it is not trivial to remove them without a loss of object information. Moreover, it is often to occur that noise have similar characteristics to objects of interest. Thus, it is quite problematic to a computer to automatically distinguish them by considering their features.

One of the most powerful tools for noise suppression is image smoothing (also known as lowpass filtering). Image smoothing suppresses the noise by attenuating its signal which makes its intensity roughly consistent with those of its nearest neighbors. Unfortunately, in many cases, i) polluted objects are much clearer and sharper than spines (in *Scenedesmus*, *Xanthidium*, and *Staurastrum* genera) and flagellums (in *Euglena* and *Phacus* genera) of algae; and ii) a thick gelatinous coat of algae is sharper than a true algae boundary. If we perform a high degree of noise suppression in order to remove polluted objects and a gelatinous coat, this usually removes spines, flagellums, and internode contours of these algae. On the other hand, if we perform a low degree of noise suppression, the detected boundary of algae body often distorts and lies further away from the true boundary of the algae (due to touching polluted objects and a thick gelatinous coat of algae) .

This situation causes a serious problem to classifying algae in *Anabaena*, *Oscillatoria* and *Pleurotaenium* genera. The algae in these genera have similar rod shape. The main difference between their shapes is that algae in *Oscillatoria* and *Pleurotaenium* genera have smooth boundary, while algae in *Anabaena* genus have ripple along its boundary. If we perform insufficient image smoothing, the ripple along the boundary of algae in *Anabaena* genus disappears (due to gelatinous coat), and the smooth boundary of the algae in *Oscillatoria* and *Pleurotaenium* genera becomes ripple (due to small touching polluted objects).

In order to handle this difficulty, we thus classify algae images into two groups: a rod shape and a non-rod shape groups. Algae in *Anabaena*, *Oscillatoria*, *Closterium*, *Pleurotaenium* and *Euglena* are classified into the rod shape group, while the rest are classified into the non-rod shape group. Algae



(a) (b) (c) (d) (e) (f) (g)

Figure 2: An example of our image segmentation method: (a) Original image (b) Gradient magnitude image (c) Edge image (d) Filling the boundary gaps (e) Filling the holes (f) Eroding shape (g) Removing unwanted particles

images in the rod shape group require high degree of image smoothing so that it can extract the contour as close as possible to the true algae boundary. The non-rod shape group requires less degree of image smoothing in order to preserve their spines or flagellums. Thus, in a segmentation process, the two groups of algae image are treated separately. In this work we propose *a single-resolution edge detection* for segmenting images in the non-rod shape group (because they require small amount of image smoothing) and propose *a multi-resolution edge detection* to handle with images in the rod shape group (because they require higher amount of image smoothing).

3.2.1 A single-resolution edge detection method

This segmentation method is designed to be applied to algae images in a non-rod shape group that we have to preserve spines or flagellums. Thus, we skip a smoothing process and start the algorithm with Sobel edge detection on a grayscale image. The resulting gradient magnitude image is then put to the Canny edge detection to produce an edge image. The edge image is a binary image where the 1 pixels indicate edge pixels and the 0 pixels indicate non-edge pixels. In the Canny edge detection [2] process, we use a small value of smoothing parameter, namely, $\sqrt{2}$ in order to preserve as much as possible edges of algae body. At this step, the area inside the algae body may possibly be full of holes and the algae boundary is not always connected. We fix this by applying morphological operators to the edge image. Gaps along the algae boundary are filled by using a dilation operator with a bar-shaped structuring element (SE) of size 2 pixels. The operation is performed in both vertical and horizontal directions. After the boundary of algae are connected, a hole-filling operator [16] is performed in order to fill holes in the algae body. The final step is to erode the shape of algae body back to its original size by using an erosion operator with a bar-shaped SE of size 2 pixels in both vertical and horizontal directions (The shape of algae body has been dilated in the process of filling gaps along the algae boundary).

In practice, algae images are often polluted by unwanted objects or illumination artefacts. It is general that segmentation results obtained from the above steps usually contain isolated pixels/regions around a shape and a background. Therefore, a postprocessing process usually needs to be performed to eliminate those isolated pixels and small regions from a segmentation result. This can be done by applying morphological erosion with a diamond-shaped SE of size 3×3 pixels. The operator is performed in both vertical and horizontal directions. An example of image segmentation using a single-resolution edge detection method is shown in Figure 2.

3.2.2 A multi-resolution edge detection method

In a multi-resolution edge detection method, some additional steps of image smoothing are added in addition to a normal process of a single-resolution edge detection method. A multi-resolution edge

detection method has two major steps: an initial step and a refinement step, as followings:

An initial step: Similar to a single-resolution edge detection method, the first step of this method starts with Sobel edge detection and followed by the Canny edge detection. However, this time a large value of smoothing parameter of the Canny edge detection is used (i.e. $\sqrt{10}$). A large value of smoothing parameter is used because in an initial step we only need to roughly estimate the boundary of the algae in an image. The true boundary will be detected in the next refinement step.

A refinement step: After separating the algae body from a background, only the background of the image will be heavily smoothed by using a Gaussian lowpass filter of size 20×20 pixels with sigma equal to 0.5 in order to suppress all unwanted objects and illumination artefacts in the image background. The foreground of the image (i.e. the algae body) is left unsmoothed because we want to preserve as much as possible its edge details.

After a smoothing process, the smoothed image will simply be segmented by the single-resolution edge detection method described above. A multi-resolution edge detection method derives its name from the fact that edge detection is performed on an image with different smoothing resolutions of foreground and background regions.

In practice it does not know beforehand whether a new input image belongs to a rod shape or a non-rod shape groups. Thus, we generate the SMO classifiers in advance by using a single-resolution edge detection method. The only three shape features, namely ratio of major and minor axis length, convex area, and ratio of region area and area of its bounding box are used as shape features in a classification process. Based on our preliminary experimental results, these features are sufficient for classifying rod shaped algae from non-rod shaped algae. Additionally, these features are simple and fast computation. The overall image segmentation method is summarized as following steps:

1. Performing image segmentation using a single-resolution edge detection.
2. Extracting three shape features: Ratio of major and minor axis length, convex area, and ratio of region area and area of its bounding box.
3. Classifying a new image into either rod-shaped or non-rod shaped groups using SMO classifier based on the three shape features.
4. If the new image is classified as a rod-shaped alga, the image is re-segmented by using a multi-resolution edge detection method to produce a new segmentation result. Otherwise, the segmentation result computed in the first step is used.

3.3 Feature Extraction

Object classification is a process of classify the observations into several genera. It performs by making decisions on the basis of several features measured from an object. Several object features have been studied in the literature and successfully used in practice, for example, color, shape, texture, and corner. However, color and corner features do not seem to be very useful in our problem (as discussed earlier). Hence, only shape and texture features are considered in this work. Shape features work effectively in our problem since algae in each individual genus typically have their own unique shape. However, it is not uncommon that two or more genera have similar shapes. In this case, texture features play an important role in a classification task. We note here that algae images in our data set have different sizes and algae in each image have different directions or rotations. Thus, it is better to use algae features that are invariant under scaling and rotation. In this work, three shape descriptors: Fourier descriptors,

moment invariant, and shape measures are considered (will be described later in this section). We also propose a new texture feature computation method based on the texture descriptors proposed by Haralick et al. [7].

We also note here that all object descriptors proposed in the literature have their own strengths and weaknesses. Each descriptor is computed from different basis of object information and no object feature can work well for all problems. Thus, we propose to apply a combination of multiple shape descriptors and texture feature in a classification process. In addition, using combination of multiple features is also beneficial for compensating segmentation errors made by a segmentation algorithm. Objects (algae bodies) extracted in a segmentation step normally contain some segmentation errors. Calculating shape and texture features from them can yield inaccurate object descriptors. Fortunately, different descriptors are computed from different object information, thus, different descriptor errors are produced. Hence, using multiple features can compensate the errors made by each other, resulting in improvement of the classification accuracy.

3.3.1 Fourier Descriptors

Fourier descriptors are successfully used in shape discrimination and shape analysis [12]. Because their nice properties, such as simple derivation, simple normalization, and robustness to noise. They have been extensively applied in many areas [3, 9]. Traditionally, the Fourier descriptors are not invariant to scaling and the starting point. We have to normalize them so that they are invariant under these conditions [22].

3.3.2 Moment Invariants

Moments invariants have been extensively used to characterize shape of objects in a variety of applications. Moments invariants are region-based shape descriptors and derived from information of all pixels in a shape region. When relatively large amount of noise is present in an image, this approach is more accurate than contour-based approach because it takes much more image pixels into account. In this work we use the Hu's seven moment invariants [8] that have the desirable properties of being invariant under image translation, scaling, and rotation.

3.3.3 Shape Measures

Shape measures measure the properties of a shape region in a binary image. Ten measurements of shape properties are investigated, namely, area, major axis length, minor axis length, major axis length and minor axis length ratio, eccentricity, convex area, diameter, solidity, extent, and perimeter. These measures are invariant under rotation, but not under image scaling. Thus, we have to normalize the size of shapes before compute the shape features. We normalize the size of shapes regarding to the height of the bounding box of shapes. All shapes are resized into the same height of bounding boxes while preserving aspect ratio of the shape. In this work we set the height of the bounding box to 200 pixels.

3.3.4 Texture Features

Texture is one of the most important characteristics used for identifying objects. In this work we used texture descriptors proposed by Haralick et al. [7]. In order to compute the texture descriptors, a gray-level co-occurrence matrix (GLCM) is first computed. This matrix is computed using grey-level values within an object region. We extract grey-level values within the object region by using a segmentation image as a mask. Following the work [15], we improve the GLCM by eliminating paired relationships

with background components appeared in the first row and column of GLCM. Finally, Haralick's thirteen texture descriptors are computed based on this matrix (For more detail of these descriptors we refer the reader to the original work).

The proposed texture descriptor method

Unfortunately, some of our algae images have unclear texture. A GLCM computed by using above method may contain a substantial number of errors and results in inaccurate texture descriptors. This problem is critical when texture is the only feature that can be used to discriminate algae of different genera with similar shape features. In this work we propose a new method for computing texture features from a blurred texture object. The proposed method compute new texture descriptors by averaging the texture descriptors extracted from the input image with different levels of edge enhancement. Edge enhancement is performed in order to highlight blurred texture; however, it also highlights noise in an object. As a result, the texture descriptors computed from an input image with different levels of enhancement are not accurate. We then propose to average them to produce a new and more accurate texture features because averaging is a good concept for effectively suppressing noise in data. In this work, three levels of edge enhancement are performed in an input image. For each level we compute the Haralick's thirteen texture descriptors. The final texture descriptors are the average of the three sets of descriptors. Our preliminary experimental results indicate that using the new (averaging) version of texture descriptors yield better classification accuracy than using the original (unenhanced) version of texture descriptors. In this work, an unsharp masking technique is used for edge enhancement.

4 Experiments

4.1 The Classifiers used in the Experiments

In this work we propose to use SMO classifier in a classification process. Our proposed method is evaluated in comparison with three effective, well-known classifiers, namely, multilayer perceptron (MLP) [19], Bagging [19], and J48 decision tree. All classifiers were applied using the same set of object features. The parameter values of each classifier were tuned in such a way that the highest average of recognition accuracy was obtained. For each set of feature combinations, the same parameter settings of each classifier are applied. For MLP classifiers, a number of hidden nodes are dynamically determined by a number of feature dimensions. The work [6] suggested that we should increase a number of hidden nodes when a number of feature dimensions increases. For SMO classifiers, we set a value of complexity parameter to 4. For Bagging classifiers, a number of training iterations was set to 10. For the J48 decision tree, a default parameter setting set by WEKA is used. All classifiers used in the experiments are provided by WEKA [6]. The dataset is divided by biologists into 540 training images (45 images per class) and 180 test images (15 images per class).

4.2 Tuning the Parameters of Feature Descriptors

The accuracy of Fourier descriptors depends heavily on a number of sampling points. We empirically tested the performance of Fourier descriptors by varying a number of sampling points. The experimental results show that an appropriate number of sampling points is 32 points, which is sufficient for describing fine details of shape contours with reasonable computation time. The accuracy of GLCM-based texture descriptors depends on a number of gray levels used for computing the GLCM. We empirically tested the performance of GLCM-based texture descriptors by vary the number of gray levels. The experimental results suggest that the most effective number of gray-levels is 256 gray levels. The total number of features used in the experiments are summarized in the 7th column of Table 2.

Approach	No.	Image features				No. of features	Accuracy (%)			
		Moment	Shape	Fourier	GLCM		MLP	SMO	Bagging	J48
Single image features	1	✓				7	66.11	37.22	74.44	72.22
	2		✓			10	84.44	81.67	81.67	79.44
	3			✓		30	80.00	82.22	83.89	71.11
	4				✓	13	23.89	33.33	26.11	21.67
Combination of multiple image features	5	✓	✓			17	86.11	82.22	87.22	80.00
	6	✓		✓		37	81.67	90.00	88.89	82.22
	7	✓			✓	20	52.22	58.33	79.44	74.44
	8		✓	✓		40	85.56	92.22	85.56	75.56
	9		✓		✓	23	84.44	91.67	81.67	75.56
	10			✓	✓	43	85.00	91.11	82.22	70.00
	11	✓	✓	✓		47	85.56	91.67	90.00	81.11
	12	✓	✓		✓	30	86.67	92.78	89.44	75.56
	13	✓		✓	✓	50	88.33	93.89	87.22	82.78
	14		✓	✓	✓	53	90.00	96.11	86.67	76.67
	15	✓	✓	✓	✓	60	90.00	**97.22**	89.44	75.56

Table 2: Classification results of four classifiers on 180 images of 12 genera of microalgae with different sets of image features

Genus	1	2	3	4	5	6	7	8	9	10	11	12	Accuracy
Anabaena	15	0	0	0	0	0	0	0	0	0	0	0	100.00%
Oscillatoria	0	13	0	0	0	0	0	0	0	2	0	0	86.67%
Microcystis	0	0	14	0	0	0	0	1	0	0	0	0	93.33%
Scenedesmus	0	0	0	14	0	0	0	0	1	0	0	0	93.33%
Pediastrum	0	0	0	0	15	0	0	0	0	0	0	0	100.00%
Cosmarium	0	0	0	0	0	15	0	0	0	0	0	0	100.00%
Closterium	0	0	0	0	0	0	15	0	0	0	0	0	100.00%
Xanthidium	0	0	0	0	0	0	0	15	0	0	0	0	100.00%
Staurastrum	0	0	0	0	0	0	0	0	15	0	0	0	100.00%
Pleurotaenium	0	0	0	0	0	0	0	0	0	15	0	0	100.00%
Euglena	0	0	0	0	0	0	1	0	0	0	14	0	93.33%
Phacus	0	0	0	0	0	0	0	0	0	0	0	15	100.00%
Average Accuracy													97.22%

Table 3: Confusion matrix of SMO classifier using the combination of four image features (dataset 15)

4.3 Classification Performance

We conduct a series of experiments and discuss their results in two approaches. The first approach is using a single object features and the second approach is using a combination of multiple object features. The accuracy of different approaches are reported in Table 2. A single object feature approach is reported from dataset number 1 to 4, while a combination approach is reported from dataset number 5 to 15.

4.3.1 The Performance of using a Single Object Features

In a single object feature approach, texture features (dataset number 4 in Table 2) give the lowest classification accuracy in comparison with the three shape descriptors. This is because geometric information of algae shapes is more discriminative than texture information of algae. As we discussed earlier, an intensity variation of algae body can be largely varied due to illumination adjustment in an imaging process or due to environmental conditions in which algae are growing. Thus, using texture feature alone is hard to achieve good classification accuracy. Nevertheless, they are still useful for discriminating algae

of different genera having similar shape features from each other.

Among the three shape descriptors, the moment invariants (dataset number 1 in Table 2) give the lowest classification accuracy. Even though the moment invariants have been proved to be invariant under image scaling and rotation, the proof was made under the assumption of continuous image functions and noise-free. In practice, images are discrete and prone to noise. Therefore, errors are inevitably presented during computation of the moment invariants, resulting in low classification accuracy. However, the experimental results show that even though the moment invariants seem to have small discrimination ability, it still contributes to a classification of algae in a feature combination approach. Finally, Shape measures and Fourier descriptors perform comparable performance. Although algae boundaries contain some segmentation errors, both descriptors have shown their robustness to segmentation errors.

4.3.2 The Performance of using a Combination of Multiple Features

From Table 2, we can notice that the accuracy of using multiple features is mostly higher than the accuracy of using a single feature for all classifiers (except for J48 decision tree). The main observations of these experimental scenarios are summarized as following: i) Using a combination of only shape features (dataset number 11) is not sufficient for discriminating algae with similar shape features that belong to different genera correctly. We can conclude here that using only shape features is not able to achieve desirable classification accuracy, and this indicates the need of texture features; ii) The highest classification accuracy we can achieve is 97.22% provided by SMO classifiers with a combination of all feature descriptors (dataset number 15). This indicates the usefulness of texture features that help classifiers correctly classified algae with similar shapes into different genera; iii) The classification accuracy of SMO classifier dropped to 96.11% when the moment invariants were not used in a training process (dataset number 14). This demonstrates the usefulness of the moment invariants. The moment invariants can be used to compensate errors or ambiguities of Fourier and shape measurement descriptors.

Table 3 shows a confusion matrix of SMO classifier with a combination of four descriptors (dataset number 15). We achieve 100% classification accuracy in most classes, except for *Oscillatoria*, *Microcystis*, *Scenedesmus*, and *Euglena*. Two algae of *Oscillatoria* genus are misclassified to *Pleurotaenium* genus because they have almost the same shape features and texture of these two algae is not clear. One alga of *Microcystis* genus is misclassified to *Xanthidium* genus because the shape of this alga is largely different from the algae in its genus. One alga of *Scenedesmus* is misclassified to *Staurastrum* genus because both algae have similar pattern of their spines. Finally, one alga of *Euglena* is misclassified to *Closterium* genus because its shape is more similar to shapes of algae in *Closterium* genus than shapes of algae in its own genus. Moreover, the texture of this alga in an image is substantially unclear. Thus, texture descriptors cannot help in this case.

5 Conclusion and Future Work

This paper presented an automated microalgae recognition method for microscopic images. A new image segmentation method for separating algae from an image background was proposed. Our proposed segmentation method can deal with several segmentation difficulties such as unclear algae boundary, transparent appearances of spines and flagellums of algae, and touching polluted objects. The causes of the unclear boundary can occur during image acquisition process and can be caused by the characteristic of algae themselves. The spines and flagellums of algae look transparent when they are captured by a camera. Often, these parts of algae in an image are much more blurred than extraneous objects polluted in an image. Our new segmentation method based on single- and multi-resolution edge detection can handle this situation well. Moreover, we proposed a new method for computing texture descriptors from

blurry texture objects. Experimental results demonstrate the effectiveness of our proposed segmentation method and our proposed texture descriptor computation method.

Even though our first results are promising, many tasks remain for future work to improve the performance of the current system. For example, it is necessary to evaluate the performance of the proposed method on a larger set of image data. The feature selection methods should be investigated and applied in order to reduce a number of features used in a classification process. Furthermore, it is essential for a classifier to have a rejection mechanism to reject unknown algae or extraneous objects that are unknown to the trained classifier. Rejecting the unknown objects would make the recognition system more reasonable than classifying them into any incorrect class.

Acknowledgements

This work is supported by Kasetsart University Research and Development Institute under Grant No.154.56.

References

[1] S.Z. Boujelbene, D. Ben Ayed Mezghani, and N. Ellouze. Vowel phoneme classification using SMO algorithm for training support vector machines. In *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*, pages 1–5, April 2008.

[2] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679–698, November 1986.

[3] S. Conseil, S. Bourennane, and L. Martin. Comparison of Fourier descriptors and Hu moments for hand posture recognition. In *European Signal Processing Conference (EUSIPCO)*, 2007.

[4] Ria Rodette G. de la Cruz, Trizia Roby-Ann C. Roque, John Daryl G. Rosas, Charles Vincent M. Vera Cruz, Macario O. Cordel, Joel P. Ilao, Adrian Paul J. Rabe, and J.Parungao Petronilo. SMO-based system for identifying common lung conditions using histogram. In *Medical Information and Communication Technology (ISMICT), 2013 7th International Symposium on*, pages 112–116, March 2013.

[5] S. N. Deepa and B.A. Devi. Neural networks and SMO based classification for brain tumor. In *Information and Communication Technologies (WICT), 2011 World Congress on*, pages 1032–1037, December 2011.

[6] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.

[7] R.M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(6):610–621, November 1973.

[8] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179–187, February 1962.

[9] Chung-Lin Huang and Dai-Hwa Huang. A content-based image retrieval system. *Image and Vision Computing*, 16(3):149 – 163, 1998.

[10] Tong Luo, Kurt Kramer, Dmitry Goldgof, Lawrence O. Hall, and Scott Samson. Learning to recognize plankton. In *in Proc. IEEE Int. Conf. Systems*, pages 888–893, 2003.

[11] Mogeeb AA Mosleh, Hayat Manssor, Sorayya Malek, Pozi Milow, and Aishah Salleh. A preliminary study on automated freshwater algae recognition and classification system. *BMC Bioinformatics*, 13(17):1–13, 2012.

[12] E. Persoon and King-Sun Fu. Shape discrimination using Fourier descriptors. *Systems, Man and Cybernetics, IEEE Transactions on*, 7(3):170–179, March 1977.

[13] John C. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998.

[14] Santi Sarabol, Srunya Vajrodaya, Chatchai Ngernsaengsaruay, and Nuttha Sanevas. Diversity of algae in Khlong Kamphuan watershed, Kamphuan sub district region, Suk Samran district, Ranong province. *Thai Journal of Botany*, 2 (Special Issue):33–45, 2010.

[15] Olfa Ben Sassi, Lamia Sellami, Mohamed Ben Slima, Khalil Chtourou, and Ahmed Ben Hamida. Improved spatial gray level dependence matrices for texture analysis. *International Journal of Computer Science & Information Technology*, 4:209–219, 2012.

[16] Pierre Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2 edition, 2003.

[17] Pairin Sudthang. Diversity of algae and water quality assessment in sediment areas at Bueng Boraphet. Master's thesis, Department of Botany, Kasetsart University, 2011.

[18] Pairin Sudthang, Srunya Vajrodaya, Srisom Suwanwong, and Nuttha Sanevas. Diversity of algae in Bueng Boraphet, Nakhon Sawan province. *Thai Journal of Botany*, 2 (Special Issue):21–31, 2010.

[19] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.

[20] Jiang Tao, Wang Cheng, Wang Boliang, Xie Jiezhen, Jiao Nianzhi, and Luo Tingwei. Real-time red tide algae recognition using SVM and SVDD. In *Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on*, volume 1, pages 602–606, October 2010.

[21] Stefan U. Thiel, Ron J. Wiltshire, and Lance J. Davies. Automated object recognition of blue-green algae for measuring water quality – A preliminary study. *Water Research*, 29(10):2398 – 2404, 1995.

[22] Timothy P. Wallace and Paul A. Wintz. An efficient three-dimensional aircraft recognition algorithm using normalized Fourier descriptors. *Computer Graphics and Image Processing*, 13(2):99 – 126, 1980.