A New Shape Descriptor and Segmentation Algorithm for Automated Classifying of Multiple-Morphological Filamentous Algae^{*}

Saowanee Iamsiri^{1[0000]}, Nuttha Sanevas^{2[1111]}, Chakrit Watcharopas^{1[2222]}, and Pakaket Wattuya^{1[3333]}

¹ Department of Computer Science, Kasetsart University, Bangkok, Thailand saowanee.ai@ku.th, chakrit.w@ku.th, pakaket.w@ku.th http://cs.sci.ku.ac.th/

² Department of Botany, Kasetsart University, Bangkok, Thailand fscintsv@ku.ac.th http://botany.sci.ku.ac.th/

Abstract. In our previous work on automated microalgae classification system we proposed the multi-resolution image segmentation that can handle well with unclear boundary of algae bodies and noisy background, since an image segmentation is the most important preprocessing step in object classification and recognition. The previously proposed approach was able to classify twelve genera of microalgae successfully; however, when we extended it to work with new genera of filamentous algae, new challenging problems were encountered. These difficulties arise due to a variety of the forms of filamentous algae, which complicates both image segmentation and classification processes, resulting in substantial degradation of classification accuracy. Thus, in this work we propose a modified version of our multi-resolution segmentation algorithm by combining them in such a way that the strengths of both algorithms complement each other's weaknesses. We also propose a new skeleton-based shape descriptor to alleviate an ambiguity caused by multiple morphologies of filamentous forms of algae in classification process. Effectiveness of the two proposed approaches are evaluated on five genera of filamentous microalgae. SMO is used as a classifier. Experimental result of 91.30% classification accuracy demonstrates a significant improvement of our proposed approaches.

Keywords: Microalgae Image Classification \cdot Image Segmentation \cdot Shape Descriptor \cdot Skeleton.

1 Introduction

Microalgae are important aquatic life forms and live in most aquatic ecosystems. Because of responding strongly to environmental changes, they have long been

^{*} Supported by Kasetsart University Research and Development Institute under Grant No.146.58.

2 S. Iamsiri et al.

used as biological indicators for an assessment of water quality [21]. In recent studies of diversity of microalgae [14, 13] the researchers found that some species of microalgae can tolerate to contaminated water resources and can be used to phytoremediate such contaminated sites. Thus, in water resource management, a regular assessment of water conditions allows early warning of deteriorating conditions that can be toxic to both humans and animals. Rapid, accurate identification of microalgae is therefore one of the most important issues in water resource management. While the need of identification of algae increases, only a few biological experts with taxonomy competence are available. Moreover, a task of identifying the species of algae is also time consuming. Thus, an automated algae image classification is essential.

An image segmentation is a process of separating objects of interest from an image background, e.g., detecting algae bodies. It is the most important preprocessing step in an automated image classification task, especially in algae image classification. Since a shape of algae is a key characteristic used for taxonomical identification of its genus, we expect the segmentation algorithm capable of producing segmented results with property that generates boundaries of algae bodies lying as close as possible to the true edges of algae bodies, and preserves important morphological features of algae as many as possible. Completeness of morphological features of the extracted shape will contribute to high quality of shape descriptors, which finally contributes to high classification accuracy. In our previous work on automated microalgae image classification [12] we proposed a segmentation algorithm (i.e. the multi-resolution edge detection) that can efficiently deal with problems of unclear boundary of algae bodies and noisy background. The proposed algorithm worked well with twelve genera (both nonfilamentous and filamentous forms) of microalgae. Unfortunately, when applied to new genera of filamentous algae, the algorithm produced unexpected results.

Five genera of microalgae studied in this work are ones of the most commonly found in water resources of Thailand (as shown in Fig. 1). The two genera, i.e. *Anabaena* and *Oscillatoria*, are from the previous work, and the three new genera, i.e. *Spirogyra*, *Spirulina*, and *Anabaenopsis*, are newly collected. All five genera are filamentous algae. In biology, the filamentous forms of these algae are called trichomes. Trichomes can be found in many characteristics. In this work we categorize them into three groups according to their silhouette shape (as shown in Fig. 2) as following:

- 1. Smoothed boundary: trichomes in this group are composed of cylindrical cells. Thus, they have a smoothed contour along their elongate body. Microalgae in this group are *Oscillatoria* and *Spirogyra* (see Fig. 2b).
- 2. Crenate boundary: trichomes in this group are composed of spherical cells. So, they have notches at joints of spherical cells along their contours, and hence, the contour appears crenate along their length. Microalgae in this group is *Anabaena* (see Fig. 2a).
- 3. Spirally coiled boundary: trichomes in this group can be short and regularly spirally coiled like *Anabaenopsis* or irregularly spirally coiled, such as (e.g. *Spirulina* (see Fig. 2d, 2e).



(a) Crenate (b) Smoothed (c) Smoothed (d) S.Coiled (e) S.Coiled

Fig. 2: Silhouette of algae in Fig. 1.

The main purpose of inventing the multi-resolution edge detection is that we want to extract algae body from a background, where its boundary lies as close as possible to the true edges of algae, so that the shape features of algae body will be preserved as many as possible. However, when facing with new microalgae of filamentous genera, the detected boundary of algae seems not quite fit enough to efficiently capture small details of boundary in each group. The following examples show the difficulties we faced when the detected boundary of algae lies further away from the true edges of algae: i) when size of Anabaena in an input image is quite small or the trichomes are composed tightly of small spherical cells (see Fig. 3a), notches along its crenate boundary in the segmentation result are shallow. As a result, its boundary appears more like smoothed contour than crenate contour. Consequently, Anabaena is more likely to be misclassified to Spirogyra or Oscillatoria, and ii) when segmenting a tiny spirally coiled algae (Spirulina), curls occurring along a spirally coiled boundary will appear like notches (see Fig. 3b), resulting in ambiguous classification between the two genera (Spirulina and Anabaena). Finally iii), the multi-resolution edge detection was not designed to handle trichomes of spirally coiled form that curl in a circle (or a ring shape), and hence segmentation results of algae of this morphology wrongly turn into a coccoid form (see Fig. 3c). Thus, in this work we propose a new modified version of the multi-resolution edge detection so that segmentation results of algae with ring shape remains a ring shape, and notches along crenate boundary and curls of spirally coiled boundary are distinct. The experimental results show that our new segmentation algorithm does benefit a classification process by bringing the differences of the boundaries in the three groups clearer.

In addition, in order to alleviate an ambiguity of classifying ambiguous multiple morphologies of the trichomes forms, we propose a new shape descriptor computed based on a skeleton of a shape. The experimental results show that by using our new shape descriptor, the classifier can efficiently cope with several cases of ambiguity, for example, i) discriminate shapes with crenate boundary with tiny notches from smoothed boundary shapes, and ii) discriminate shape with spirally coiled boundary with tiny curls from shapes of crenate boundary.

3



Fig. 3: Incorrect segmentation results of the old approach [12].

2 Related Works

Microalgae segmentation as well as automated algae classification are a very difficult task due to various factors: not only a wide variation of characteristics of algae in nature, but also the poor conditions of microscopic images (e.g. low contrast due to non-uniform illumination conditions, strong background noise arising from lights scattered, presence of various extraneous objects in the background). Therefore, algae image classification and its associated problems become challenging for applications of computer vision [16]. Various image processing and machine learning approaches have been proposed to tackle these problems.

Edge-based segmentation approach is one of the most popular approaches that widely uses in microalgae image segmentation because of its robustness to noise. Li et. al. [7] proposed to use Sobel edge detection for semi-automated microorganism segmentation. Since Sobel edge detection often produces discontinuous boundaries, Santhi et. al. [16] and Promdaen et. al. [12] proposed to use Canny edge detection on a result of Sobel edge detection in order to improve segmentation results. A Canny edge detection was also used by [9,8]. Because most of edge detection methods naturally requires postprocessing for linking discontinuities of edges to form closed contours of objects, deformable model approaches become more attractive since they produce results as a closed contour. Giraldo et. al. [5] proposed to used active contour, while Borges et. al. [2] proposed to use level set for segmenting microalgae images. However, edge-based active contours are sensitive to noise and may omit minute features and blurry boundaries during energy minimizing. Therefore, this approach is not suitable for our problem at hand, in which input images may contain both blurry-boundary objects and noisy background. Besides edge-based approach, Zheng et.al. [24] proposed an interesting intensity-based image segmentation approach, namely GSDAM. GSDAM separates an object from a background by considering the connectivity and directionality of lineal structures of intensity information. It can handle well with low contrast as well as noisy background. Unfortunately it is particularly designed for segmenting objects with thin and long shape.

Feature extraction and description are also of important for automated algae classification. Shape descriptors are inevitable primary features for this task, while size and texture depend on the nature of the problem. For example, the works of [7,9,16,22,1] could use size of objects as one of key features because their input images came from a single source, so size of algae in an image are unvarying. In contrary to our work where input images come from several sources with unknown imaging conditions, the size of algae are varying in a wide range, and thus using size as a key feature merely complicates the problem. Com-

monly used shape descriptors are such as geometric shape descriptors [7, 9, 16, 22], Fourier descriptors [12], and moment invariant [5, 22]. A major drawback of moment invariants is that higher-order moments are very sensitive to noise. Texture is also one of the most important characteristics used for identifying algae. Most commonly used texture descriptors in this area of problems are Haralick [5, 8] and local binary pattern (LBP) [5, 22]. Because of the way LBP captures texture information (i.e. considering difference of pixel intensity, while neglecting magnitude of intensity), its major drawback is a limitation of capturing the texture information.

For a classification task the most commonly used classifiers are support vector machine [5, 7, 12, 22] and an artificial neural network [5, 8, 9, 16]. For more review about microorganism classification we refer to an interesting work of Li et.al. [7]. They provide a comprehensive review of 240 related works on the applications of content-based microscopic image analysis in microorganism classification domains.

3 A New Segmentation Algorithm

In our previous work [12] we proposed two segmentation algorithms, namely the single-resolution edge detection and the multi-resolution edge detection. The single-resolution edge detection was purposely designed for segmenting microalgae of coccoid form, while the multi-resolution edge detection was purposely designed for segmenting microalgae of filamentous and tube-like forms. The rationale behind our idea of treating the two groups of algae differently is that algae of coccoid form require a small degree of image smoothing in an edge detection process in order to preserve their spines or flagella, while filamentous and tube-like algae require high degree of image smoothing so that an edge detector can extract an algae body whose boundary lies as close as possible to the true algae edges. In this work we still adhere to this idea and attempt to extract an algae body whose boundary fits with the true boundary of algae as much as possible. Since the multi-resolution edge detection already performs well with filamentous algae with smoothed boundary and has strong ability to deal with a noisy background, our plan is to combine a segmentation result of the multi-resolution edge detection with one of a new proposed algorithm. The new segmentation algorithm now combines the previous abilities with one that can handle a wide variety of morphology of filamentous forms. The new algorithm is called a generalized multi-resolution segmentation algorithm, since it can deal with a wide range of filamentous forms. The new algorithm is composed of four main steps. The details of each step are described as follow.

Step 1: Preprocessing

A preprocessing is a process of preparing an input image to have suitable properties for successive processes, e.g. image segmentation and feature extraction. The preprocessing that performs in this work is similar to one described in the previous work [12]. Only a process of converting an RGB input image into a grayscale image is reconsidered. In this work we found that edges of algae in a blue channel of input RGB image appear more distinguishable from a background than those

6 S. Iamsiri et al.

in a grayscale image obtained by a color-to-gray transformation. Thus, in this work we propose to use a blue channel of input image as a grayscale version of input image, instead of a grayscale image obtained by a color-to-gray transformation. And since edges of algae play an important role in a segmentation process, we perform edge enhancement on a grayscale image using the Contrast-Limited Adaptive Histogram Equalization (CLAHE) [11] to improve its acutance.

Step 2: Foreground/background segmentation

The objective of this step is to extract algae body from an image background in a way that the contours of algae bodies lie as close as possible to the true edges of algae bodies. To achieve this we hence decide to perform following.

- 1. We use a thresholding method instead of edge detection. The adaptive thresholding method is applied to a grayscale image, transforming the grayscale image into a binary image, where regions of algae bodies are assigned as a foreground (pixels with value 1), leaving background regions to remain as a background (pixels with value 0). After thresholding, a few number of postprocessing tasks are required to improve a quality of this intermediate segmentation result.
- 2. Small noisy objects are removed using morphological opening operator (we note here that the result of this stage will be used again in Step 3).
- 3. We then can see that thresholding method has classified some parts of the algae body as background pixels, causing small holes in foreground regions. Small holes are filled (turning background pixels from 0 to 1) by performing morphological closing operator [18] with a small-size structuring element (SE). Using a small-size SE in this step is preferable to a large-size SE because we want to prevent unwanted objects from being connected to regions of algae bodies.
- 4. The multi-resolution edge detection proposed in our previous work plays an important role in this task. We utilize its ability to cope with noisy background. Unwanted objects are removed from the segmentation result by intersecting them with the segmentation result produced by the multiresolution edge detection. As a result, segmentation results with pleasantly clean background is obtained.
- 5. Finally, large holes within foreground regions must be filled. Note that these large holes do not commonly occur. Only algae with transparent gaps in their body, e.g. *Oscillatoria* and *Spirogyra*, can produce these holes. In order to fill large holes, the result from multi-resolution edge detection is required. Firstly, we would like to note that the multi-resolution edge detection is an edge-based image segmentation approach, while the new algorithm is a region-based image segmentation approach. Edge-based segmentation segments an image by considering a discontinuity of image pixels to identify edges (or boundary) of objects, while region-based segmentation considers a similarity of image pixels and groups them together into regions (or objects). Hence, the results of these two approaches are complement to each other in a way that one's imperfect result may be completed by a result of another



Fig. 4: The whole process of Step 2. Both horizontal and vertical-line SE of size 3 pixels are used in morphological operators.

[10]. In our situation an advantage of the new segmentation algorithm is the ability to produce segmentation results in which boundary of algae body lie very closely to the true algae boundary. However, this ability comes with an unavoidable drawback; it often yields an indented region of algae body when segmenting algae with transparent gaps in their body. On the other hand, an advantage of our old segmentation algorithm is that it often yields a full regions of algae body, but boundary of extracted algae body lies a bit farther away from the true algae edges. We thus utilize the advantages of both algorithms to complement their drawbacks by combining the two results together by means of morphological reconstruction, i.e. reconstruction by erosion. Reconstruction by erosion involves two images, i.e. a marker image used as a starting condition of transformation and a mask images used to constrain the transformation. So we use a segmentation result of the old algorithm as a marker image and a result of the new algorithm as a mask image. Then a process of reconstruction by erosion begins by iteratively eroding away the boundaries of foreground regions (algae bodies) in a marker image until they fit a foreground of a mask image or reach a predefined number of iterations. We finally obtain a final segmentation result in which a full region of algae body is obtained and the boundary of algae body lies very closely to the true boundary of algae.

The whole process in this step is shown in Fig. 4.

Step 3: Background removing (for a ring-like form)

In this final step only images of filamentous algae whose bodies curl in a circle (a ring-liked shape) are involved. In this stage we can see that a background region inside a ring shape (called *area* A for short) is incorrectly assigned as part of algae body. We thus have to remove this area to form a correct ring shape. This process begins with extracting an area A and then removing it from a final segmentation result obtained from the previous step. Area A can be extracted by performing an image subtraction of the final segmentation result and the intermediate result obtained from Step 2. A subtraction result (area A) is shown



Fig. 5: The whole process of Step 3. 2-pixels disk SE are used .

in Fig. 5. We then enhance its quality by performing a hole filling operation before subtracting it from the final segmentation result, in order to achieve a new correct one (i.e. a ring-shape algae).

Since only images of ring-shape filamentous algae are involved in this step, we have to discriminate algae of ring-liked forms from the others. We simply accomplish this by using a global shape measure called *Eccentricity* (e.g. the ratio of the length of major axis to the length of minor axis of the shape). A value of eccentricity is 0 if a shape is a circle; and the value of 1 indicates a shape of a line segment. Eccentricity is computed on the final segmentation result obtained from Step 2.

4 A New Skeleton-based Shape Descriptor

In general, the accuracy of classification and recognition system depends heavily on the quality of relevant object features (or descriptors) used in a learning process. Conventionally, characteristics of algae, such as size, shape, colors, and texture, are routinely used for taxonomical recognition. However, size and color features do not seem very useful for our problem at hand. In our task microscopic images received from users are produced from several imaging systems with different settings. Consequently, sizes of algae (even in the same genus) are varied in a wide range depending on a magnifying power used in an image acquisition process. Similar situation also happens with the colors of algae in microscopic image. Colors of algae not only typically vary in nature depending on their growing environment, but also vary significantly according to illumination adjustment during imaging process. Hence, in this work shape and texture characteristics are the key features used in a learning process.

Global shape measures, such as geometric features (e.g. compactness, rectangularity, etc.) are a convenient way to describe shapes and are simple to compute [23]. Unfortunately, these features are unable to capture a small difference along a boundary of two different shapes, e.g. shapes with crenate boundary with tiny notches and shapes with wavy boundary with tiny curls. A more sophisticated shape descriptor such as Fourier Shape Descriptors also requires high computational time because a high number of Fourier coefficients must be used, in order to be able to distinguish small differences along the boundary of two different shapes.

In this paper we propose a new shape descriptor that utilizes the weakness of skeleton. The rationale behind our idea is that skeleton is intrinsically sensitive to small changes in a shape's boundary. A small change in the shape's boundary can cause a large change in the skeleton [4, 19]. We hence utilize the sensitivity of skeleton to capture small differences along the boundary of two different shapes. A new skeleton-based shape descriptor is computed on a binary image obtained from a segmentation process. The boundary of algae body in a binary image needs to be smoothed (e.g. by morphological opening operator) before we compute a skeleton. Unavoidable spurs will be suppressed afterwards by using morphological erosion. Then, three quantities of skeleton-based shape descriptor that measure a wavy degree of the skeleton can be computed as following.

- 1. *Waviness*. Waviness is a ratio between the length of skeleton and the length of skeleton displacement. A value of waviness is equal to 1 if the skeleton is a straight line, and closer to 0 when the skeleton is highly wavy.
- 2. Number of Peaks. Skeletons of algae with spirally coiled shape will look like sinusoidal wave. Peaks are sinusoidal peaks or location of local maxima of the curve. We count them as peaks if the height of them are higher than a predefined threshold (e.g. 3 pixels).
- 3. Average Height of Peaks. Average height of peaks is a ratio between the total height of all peaks and the number of peaks.

5 Experiments

5.1 Algae Image Dataset

Microscopic algae images of five genera, namely Anabaena, Oscillatoria, Spirogyra, Spirulina, and Anabaenopsis, used in the experiments are collected from various sources. The main sources are the study area of Bung Borapet fresh water source, Nakhon Sawan province [20] and the study area of Khlong Kamphuan watershed, Ranong province [17] in the research projects conducted by Department of Botany, Kasetsart University, Thailand. Other sources include MetropolitanWaterworks Authority and online algae image database from the internet. The data set comprises of 300 images, 60 images per genus.

5.2 Automated Algae Image Classification Framework

An automated algae image classification proposed in the previous work [12] is based on a supervised learning approach that takes an algae image as input and classifies it into one of multiple genera. The classification system composes of three main components, namely image segmentation, feature extraction, and classifier (prediction model) training process. The system works as follows. After separating an algae body from an image background using a segmentation algorithm, multiple features (or descriptors) are computed from the algae body. Multiple features are then formed into a feature vector (one feature vector represents one input image). Finally, a training process takes a set of feature vectors as an input. For supervised learning a training process needs additional input,

10 S. Iamsiri et al.

i.e. a list of labels (names of genus) associated with each input image. A classifier will then be trained based on these sets of inputs. Once the classifier is trained it can be used to predict a genus of an unknown algae image in the future. Details of image features and a classifier used in a learning process will be next described.

5.3 Image Features

As we discussed earlier, only shape and texture features will be used in the learning process. It is interesting to note that both shape and texture features play an important role in this task. Shape features will play a key role when texture features are unavailable, for example, when a size of algae in an input image is very small or when an input image with out-of-focused algae is obtained. On the other hand, texture features will play a key role when shape features are ineffective, for example, when we have to discriminate two algae of the same shape (e.g. Oscillatoria and Spirogyra). Hence a suggested approach is to combines shape features with texture features to form more powerful discriminating features used for training a prediction model. Shape descriptors used in this work are three geometric shape features plus three quantities of our new skeleton-based shape features. The three geometric shape features [23] are Solidity, Eccentricity, and Convexity (For more details of these descriptors we refer the reader to the original work). Similar to our previous work [12], texture descriptors used in this work are the thirteen texture descriptors proposed by [6]. The thirteen descriptors are derived from a gray level co-occurrence matrix (GLCM). This matrix characterizes texture of an object (or image) by calculating the frequencies of pixel pairs with specific values and in a specified spatial relationship occurred in an image. Based on our preliminary experiments, it suggests that the most effective number of grav-levels used for computing GLCM is 256 and a direction of calculation is set to 0° . The thirteen descriptors include Angular Second, Contrast, Correlation, Sum of Squares, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, and Information Measure of Correlation 1, 2. Hence, the total number of features we used in a training process are 19 features.

5.4 Classifiers

As suggested from our previous work [12], we use the Sequential Minimal Optimization (SMO) algorithm for training a support vector classifier using scaled polynomial kernels in a classification process since it has been successfully applied in many application domains such as medical image analysis, handwritten character recognition, and speech recognition. We set a value of complexity parameter of SMO to 4. Due to a limited data sample, the k-fold cross-validation method is applied in an evaluation process. In all experiments the value of k is set to 4. So, our image dataset is randomly partitioned into 4 subsets of equal size (i.e., 15 images per subset). Of the 4 subsets, a single subset is retained as the test data used for testing the trained classifier, and the remaining 3 subsets are used as training data. The training process repeatedly runs 4 times, each time with different subset of data. The 4 classification results will be averaged to produce a single classification accuracy.

5.5 Classification Results

In this section the three experiments with three different experimental settings are reported. The first experiment is a baseline for evaluating an improvement of the new proposed approaches. The old version of algae image classification [12] is applied with new algae dataset (using the old segmentation algorithm and the old set of shape features), and we let its result be the baseline's performance. The second experiment is conducted by employing the new segmentation algorithm; however, only 16 features are used in a training process. The three features of the new skeleton-based shape descriptor are omitted. Finally, the third experiment is conducted using the new segmentation algorithm and all 19 features (including the three new features). The experimental results of the three experiments are shown in Table 1, 2, and 3 respectively.

The average classification accuracy of our old approach is 84.00% (Table 1). It can be seen that algae of the first four classes are highly misclassified over each other's classes. The main cause of these mistakes is a consequence of inaccurate image segmentation. When examining the segmentation results of *Anabaena*, we found that crenate boundaries of algae's shape barely appeared. Consequently, shape of *Anabaena* becomes more similar to those of *Oscillatoria* and *Spirogyra*. Similarly, shape of *Spirulina* with tiny curls looks more like shape of *Anabaena* than those of the self-genus. These ambiguities cause a classification task to be more complicated.

Experimental results of the second experiment are shown in Table 2. The average classification accuracy is 86.30%. Even though the average accuracy does not significantly improve from the baseline performance, the ambiguity among the three classes, i.e. Anabaena, Oscillatoria, and Spirogyra, significantly decreases. A group of algae with straight boundary (Oscillatoria and Spirogyra) is much better discriminated from a group of algae with wavy boundaries (Anabaena and Spirulina). Only within a group of algae with wavy boundaries (Anabaena, Spirulina, and Anabaenopsis) are confusedly classified to other genera. This classification result explicitly demonstrates the advantage of using efficient segmentation algorithm - unsophisticated case of ambiguity can simply be alleviated.

The third experiment demonstrates that a more sophisticated case of ambiguity (i.e. an ambiguity among a group of shapes with wavy boundaries.) can be alleviated by means of involving a discriminative shape descriptor in the learning process. Experimental results of the third experiment are shown in Table 3. The average classification accuracy is significantly improved to 91.30%. This result strongly demonstrates the effectiveness of the proposed skeleton-based shape descriptor.

5.6 Discussion

In the third experiment, fewer algae in each class are misclassified into incorrect genus. After examining the segmentation results of these samples, we found nothing wrong with the results (example of them are shown in Fig.6). We thus presume that because a number of images in our dataset is quite limited, hence

11

¹² S. Iamsiri et al.

Genus	Anabaena	Oscillatoria	Spirogyra	Spirulina	${\it Anabaen opsis}$	Accuracy
Anabaena	53	3	4	0	0	88.30%
Oscillatoria	10	42	6	2	0	70.00%
Spirogyra	5	7	48	0	0	80.00%
Spirulina	8	0	0	50	2	83.30%
Anabaen opsis	0	0	0	1	59	98.30%
	Average Accuracy				84.00%	

Table 1: A confusion matrix of the old version of algae image classification [12].

Genus	Anabaena	Oscillatoria	Spirogyra	Spirulina	Anabaenopsis	Accuracy
Anabaena	48	4	2	6	0	80.00%
Oscillatoria	5	53	2	0	0	88.30%
Spirogyra	4	4	52	0	0	86.70%
Spirulina	8	0	0	50	2	83.30%
Anabaenopsis	0	1	0	3	56	93.30%
Average Accuracy					86.30%	

Table 2: A confusion matrix of the new approach with 16 features (without the new shape descriptor).

Genus	Anabaena	Oscillatoria	Spirogyra	Spirulina	An abaen opsis	Accuracy
Anabaena	52	3	1	4	0	86.70%
Oscillatoria	4	54	2	0	0	90.00%
Spirogyra	3	2	54	1	0	90.00%
Spirulina	4	0	0	54	2	90.00%
An aba en opsis	0	0	0	0	60	100.00%
Average Accuracy						91.30%

Table 3: A confusion matrix of the new approach with the new shape descriptor.

the training data are not a good representative of algae in a particular genus. On the other hand, texture descriptors, considering as a statistical approach, used in a learning process are probably not sufficient to capture a high variation of texture features of algae in each genus. In addition, shape descriptors may not be effective enough to capture very small detail of object boundary. Finally, it is interesting to note that our input images are collected from various sources. Hence size, colors, and illumination of algae in these input images are varied in a wide range. Nevertheless, our new segmentation algorithm is capable of segmenting them successfully in most case, which indicates the effectiveness of the proposed algorithm.

6 Conclusion

This paper presented a new image segmentation algorithm called a generalized multi-resolution that provides several advantages. Firstly, it can produce segmentation results in which boundary of algae fits well with the true boundary of algae. Consequently, important morphological features of algae, even in cases of very small algae or very small detail of algae boundary, are preserved. This



Fig. 6: Example segmentation results of incorrect classification.

allows a shape description method to compute high accurate shape descriptors, resulting in high classification accuracy. Secondly, the proposed segmentation algorithm can efficiently handle multiple morphologies of filamentous forms of algae. Lastly, the proposed segmentation algorithm is simple to compute and requires low computational time. So it can be used in a real-time environment effectively. In addition, we also proposed a new skeleton-based shape descriptor that effectively alleviates an ambiguity of multiple morphologies of filamentous forms of algae in a classification process. The proposed shape descriptor is also simple to compute; thus, it can also be applied in a real-time system.

It's also worth noting that biological images are often far more difficult to be processed and recognized than images of daily-life [3]. Image features and object descriptors are of important, especially in the domain where availability of image data is limited. Learning for selecting relevant image features to achieve a good recognition accuracy is still vital [15]. Even a deep learning on limited data still requires a special treatment to improve the accuracy of detection. Image augmentation, however, may help increase the amount of relevant data for deep learning, but in our case augmenting data couldn't be beneficial since algae exhibit diverse morphology.

References

- 1. M. A. Atteya, M. A. M. Salem, D. Hegazy, and M. I. Roushdy. Image segmentation and particles classification using texture analysis method. 2016.
- V. R. P. Borges, B. Hamann, T. G. Silva, A. A. H. Vieira, and M. C. F. Oliveira. A highly accurate level set approach for segmenting green microalgae images. In SIBGRAPI Conference on Graphics, Patterns and Images, pages 87–94, 2015.
- P. Coltelli, L. Barsanti, V. Evangelista, A. M. Frassanito, and P. Gualtieri. Water monitoring: automated and real time identification and classification of algae using digital microscopy. *Environ. Sci.: Processes Impacts*, 16:2656–2665, 2014.
- 4. N. D. Cornea, D. Silver, and P. Min. Curve-skeleton properties, applications, and algorithms. *IEEE Trans Vis Comput Graph*, 13(3):530–548, 2007.
- J. Giraldo-Zuluaga, A. Salazar, G. Diez, A. Gomez, T. Martínez, J. F. Vargas, and M. Peñuela. Automatic identification of scenedesmus polymorphic microalgae from microscopic images. *Pattern Analysis and Applications*, 21(2):601–612, 2018.

- 14 S. Iamsiri et al.
- R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans. on Systems, Man, and Cybernetics*, 3(6):610–621, 1973.
- C. Li, K. Shirahama, and M. Grzegorzek. Application of content-based image analysis to environmental microorganism classification. *Biocybernetics and Biomedical Engineering*, 35(1):10 – 21, 2015.
- Q. Luo, Y. Gao, J. Luo, C. Chen, J. Liang, and C. Yang. Automatic identification of diatoms with circular shape using texture analysis. JSW, 6(3):428–435, 2011.
- M. A. A. Mosleh, H. Manssor, S. Malek, P. Milow, and A. Salleh. A preliminary study on automated freshwater algae recognition and classification system. In *BMC Bioinformatics*, 2012.
- T. Pavlidis and Y. Liow. Integrating region growing and edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(3):225–233, 1990.
- S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. H. Romeny, J. B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368, 1987.
- S. Promdaen, P. Wattuya, and N. Sanevas. Automated microalgae image classification. *Procedia Computer Science*, 29:1981–1992, 2014.
- N. Renuka, A. Sood, S. K. Ratha, R. Prasanna, and A. S. Ahluwalia. Evaluation of microalgal consortia for treatment of primary treated sewage effluent and biomass production. *Journal of Applied Phycology*, 25(5):1529–1537, 2013.
- N. Renuka, A. Sood, S. K. Ratha, R. Prasanna, and A. S. Ahluwalia. Nutrient sequestration, biomass production by microalgae and phytoremediation of sewage water. *Int. Journal of Phytoremediation*, 15(8):789–800, 2013.
- H. Rezaeilouyeh, A. Mollahosseini, and M. H. Mahoor. Microscopic medical image classification framework via deep learning and shearlet transform. *Journal of Medical Imaging*, 3 4:044501, 2016.
- N. Santhi, C. Pradeepa, P. Subashini, and S. Kalaiselvi. Automatic identification of algal community from microscopic images. *Bioinformatics and Biology Insights*, 7:BBI.S12844, 2013.
- S. Sarabol, S. Vajrodaya, C. Ngernsaengsaruay, and N. 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.
- P. Soille. Morphological image analysis: Principles and applications. Springer-Verlag New York, Inc., 2 edition, 2003.
- M. Spitzner and R. Gonzalez. Shape peeling for improved image skeleton stability. In *IEEE Int. Conf. Acoust. Speech Signal Process.*, pages 1508–1512, 2015.
- 20. P. Sudthang. Diversity of algae and water quality assessment in sediment areas at Bueng Boraphet. Master's thesis, Botany Department, Kasetsart University, 2011.
- P. Sudthang, S. Vajrodaya, S. Suwanwong, and N. Sanevas. Diversity of algae in Bueng Boraphet, Nakhon Sawan province. *Thai Journal of Botany*, 2 (Special Issue):21–31, 2010.
- 22. J. Tao, W. Cheng, W. Boliang, X. Jiezhen, J. Nianzhi, and L. Tingwei. Real-time red tide algae recognition using svm and svdd. In *IEEE Int. Conf. on Intelligent Computing and Intelligent Systems*, volume 1, pages 602–606, 2010.
- M. Yang, K. Kpalma, and J. Ronsin. A survey of shape feature extraction techniques. In P. Y. Yin, editor, *Pattern Recognition*, pages 43–90. IN-TECH, 2008.
- G. Zheng, H. Zhao, X. Sun, H. Gao, and G. Ji. Automatic setae segmentation from chaetoceros microscopic images. *Microscopy Research And Technique*, 77:684–690, 2014.