BIOLOGICAL OBJECT CLASSIFICATION AND IDENTIFICATION ON LIGHT MICROSCOPY IMAGES

Mylène Roussel * & **, X. W. Tu* and D. Fontaine*

* Unité de Recherche Associée CNRS 817, Heudiasyc, Université de Technologie de Compiègne, B.P. 649, 60200 Compiègne, France.
** Lyonnaise des Eaux - Dumez, Laboratoire d'Informatique Avancée de Compiègne, Z.A.C. de Mercières, 60200 Compiègne, France;

ABSTRACT

In this article, we present a sequence of processing which permits us to recognize and identify present objects from raw light microscopy images. An application is developed to identify toxic algae, useful in water quality management. The proposed methodology - generalizable to other biological images - is to be considered as a preprocessing for further quantitative studies. The general process is composed of different steps, each of them permitting to extract the useful information on images. Thus, from a raw image, we can locate the objects of interest. Then we deal with the different feature parameters which can be extracted from the previous process. Some of them are presented in this paper. In order to obtain parameters for identification purpose, the object modelling is also needed. The work leads to the develop of a possible classification method used to identify the present objects.

I. INTRODUCTION

This work is applied to analysis of very toxical algae, namely Cyanobacteria or Cyanophyceal[1]. They product toxical toxins. We can find these algae in summer in all lakes or water areas where the ecological system is disrupted. Health problems may be related to the development of large amounts of plant biomass in the form of algal or cyanobacterial bloom in water supplies.

These algae are too small to be seen with a monocular magnifying glass, and too large to be observed under an electronical microscope. The only way to observe them is a light microscope with great magnitude lenses. The digital images may be noisy and few contrasted. The proposed low level image processing methodology has the primitive goal to keep the maximum of information during the processes.

Since geometric shape plays a very important role for the future recognition process, the extraction of the object contours naturally becomes the goal of the low level image processing.

The second part of this work is lying on the extraction of useful/good features for an efficient classification and identification making. For a practical machine vision system, the real problem is not only feature enhancement and extraction, but also feature selection which finds important and useful features for efficient classification. The extraction concerns primitive features, as well as the selection of the most important features using domain knowledge of the objects.

The last part of the present work is the application of an automatic classifier to identify the algae. The development of a classification method is shown.

II. EXTERNAL CONTOUR EXTRACTION[2]

The segmentation of well contrasted objects is not a problem, and the success of several algorithms is proved by many applications. But if the objects are poorly contrasted, it is difficult to find a threshold which leads to a right object segmentation, and in many cases, there doesn't exist a threshold for the right segmentation of the image into isolated object regions. We try to treat the problems without simplification. We use real samples of the natural world to make the preparation. No primary preprocessing (smoothing, histogram equalization or contrast enhancement) is made on input images.

Generally, the images are noisy and few contrasted. There is a lack of uniformity in the light distribution on the observation field. The background of the preparation is homogeneous enough in grey level tone. Objects appear in general as an area with some variations in grey levels surrounded by a darker boundary region. As a conclusion, the light microscopy images present very variable photometric properties which make it more difficult to find a general method applicable to each image.

The scheme of the present work relies on a particular methodology. We try to detect objects on a light microscopy image without any knowledge about these objects, nor their presence. So the process will be general in term of object shapes, as well as in term of lens selection. The scheme in figure 2 illustrates the imbricated processes to have good enough closed contours representing external contours of algae.

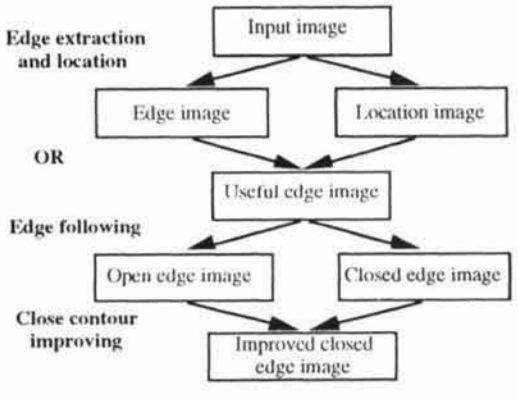
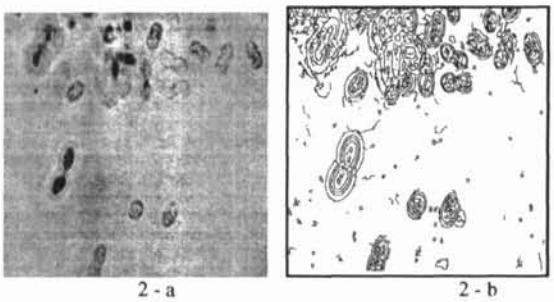


Figure 1 : Sequence of the different steps used.

Here are, in more detail, the different steps composing the general segmentation process :

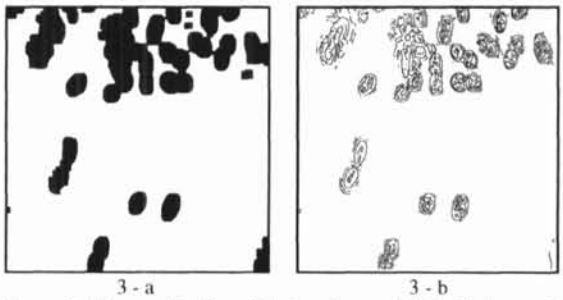
- Edge extraction : We apply recursive filters. We use an extractor derivated from Canny[3] one, named Deriche[4] filtering. Although the input images are noisy, the Deriche filter gives us a maximum of edge information without using a prior filtering.



2 - a 2 - b

Figure 2 : Example of an image of the Gloeocapsa algae taken through use of a X400 lens, and results of the Deriche recursive filtering .

- Region-based location : As the number of extracted edges is too important, we must reduce the number to keep the most useful contours. Since we have no knowledge about objects, we cannot determine to which object belongs a particular edge. A sequence of different operations is necessary to obtain a good location. We use the grey level morphological operators[5] : erosion and dilation to enhance object grey levels. Next, we apply a threshold on the approximate single Gaussian density function histogram. The resulted filtered thresholding image corresponds to a location mask image containing all objects of interest on images.



3 - a 3 - b

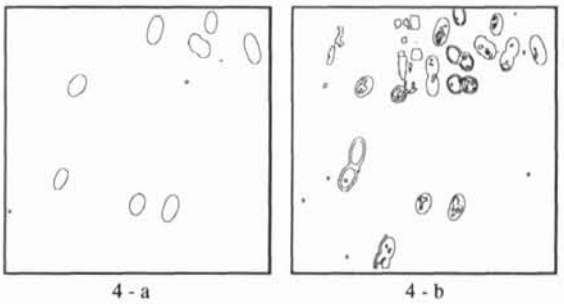
Figure 3 : The combination of the location mask (a) with the results of the Deriche filtering gives the useful edge image (b) .

- Edge following : The most useful edges are recovered by applying these location masks on the images of edges. Using the Freeman representation[6,7], we follow each edge point. We begin by all the edge points having a free extremity, keeping the detection of the closed contours for a second scan.

- Edge improving : We obtain two lists, one of the useful open edges, the other of the useful closed contours. The later should contain all the external object boundaries. In fact, some boundaries of objects are composed of one or many open edges. So, we must reconstitute the missed closed boundaries with the useful open edge image. The use of an heuristic method based on the evaluation of Euclidean distance between each open edge extremities permits us to recover more than 90 % of correct external boundaries.

- Additional filtering : Among all the close contours, we select the external alga contours. Even more, some filtering is applied to connect themselves some disconnected components and to smooth obtained contours.

The contours are now considered as the global contours of algae. We are then able to extract their characteristics.



4 - a 4 - b

Figure 4 : After the edge following, we obtain : a closed edge image(a). After the edge improving, we have the image of the improved closed contours (b).

III. FEATURE EXTRACTION

As much the low level processing is general to the segmentation of all biological objects on light microscopy images, as much this part of the work depends more on the searched objects, here algae.

1. General alga characteristics

The cyanophycea algae were received in part from the Pasteur Institute, and an other part from natural water collect (Seine water for example). A few strains was selected to demonstrate the feasibility of a classification and decision making. The criteria for this selection were : 1°/ algae of fresh water, 2°/ algae which can be found in Europe, 3°/ characteristic strains of their genus.

The cyanophycea algae are long or round in form. The algae classification is separated in few orders corresponding to that characteristic. We will placed at the genus level of the algae classification, in a family of only round algae : the chroococcal one. Five closely related strains of chroococcal family (Gloecapsa (A), Gloeothece (B), Microcystis (C), Synechococcus (D) and Synechocystis (E)) were selected.

The alga structural features, except one, can be observed by optical microscopy. As they have not a very characterized inner, the way to classify them is to get shape characteristics.

These algae have a general round shapes. Some are circular, the others will be cylindric, or in rocket or cocktail sausage form. The sizes vary from 2 micrometers to 10 micrometers. As a specification, their fission (or cellular division) never gives a filamenteous organization. The algae fission can be a criteria of identification, but it is in general difficult to use it in light microscopy. Synechococcus and Gloeothece are in rocket form. Three others are rather circular.

2. Primitive extraction

The calculation of the parameters is based partially on the Freeman representation. As we work on real natural strains, we do not use a shape proximity feature between the pattern curves of an alga model and the real algae because of the possible variations of shape. So other separative shape parameters have to be found.

Sample characteristic extraction : The global contours are clustered by mean of a Freeman chain code, and the coordinates of any points in the chain can be retrieved easily, so we can extract some basic parameters. The perimeter, the area as well as the five first order moments can be calculated from these information. The moments can be used to compute the coordinates of the geometric center of the shape. Also, they can be applied to calculate the angle of the axis of minimum inertia or

principal axis of the object.

With the next calculus, we obtain perimeter, area, geometric center of area, and principal inertia axis of each objects on the contour images. The size (2Xarea/perimeter) and the complexity (perimeter^2/4πarea), also named P2A are calculated from the two first basic parameters. From the central moments, some other parameters can be obtained as elongation and expansion. Since the objects are quite circular, a parameter based on the radius mean and variance of each objects is extracted. The ratio mean on variance gives us an indication of the circularity of the objects, different of the elongation parameter.

Elongation parameters : By the knowledge of the principal inertia axis of each object, the minor inertia axis can be found. The objects are nearly symmetrical on two orthogonal axes, and the calculation brings back to a square lattice. We obtain this new inertia axis by a $\pi/2$ rotation of the previous one.

The length and the width can be known and the rate length/width is used as another elongation parameter followed the first elongation one, but different enough. Another parameter can be also determined by the calculation of the different lengths parallel to the major inertia axis, as well as the different widths along the minor axis. We select the maximal length and width. The rate (maximal_length/maximal_width) gives a forth indication of the elongation.

On the need of other features : Let us examine now the objects. The rest of remaining parameters, although separative enough, can not be used under simple form to classify the algae. The presence of aggregate algae prohibits a good classification.

The comparison of the aggregation of two, three or four algae is not possible. An important reason is the closely shape characteristics of algae, more yet important between aggregates. Another reason is that these aggregates muddle up the space representation of extracted features. As we are not enough examples of each aggregate cases, we must separate the aggregate composing objects to obtain similar objects for comparison.

3. Fissioning alga modelling

Recall that we have no particular knowledge on the presence or not of aggregates in an object. There are two ways to determine the possibility of being on an aggregate using the contour information. First, we have two elongation parameters calculated on the same way : the ratio length/width taken on the inertia axes, and the ratio maximal_length/maximal_width along these same axis.

Constriction point extraction : On the case of a single alga in the contour, the absolute value of the difference between these two ratio will be less than 0.001. This difference takes a value greater than 0.5, for a two compound alga contour. It can achieve a value of 5. This new parameter seems to be a good indicator of the presence of an aggregate in the contour. But, using a four alga object contour gives us a difference just greater than for the single alga case. A 0.05 order value is obtained. So, this indicator is not sufficient. We must find another one.

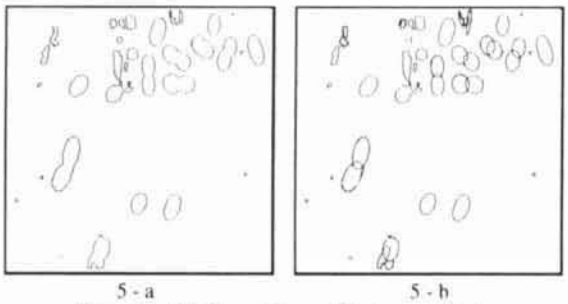
Rather than calculate another similar aggregate indicator, we prefer work on all the contour points. The described algorithm consists in finding the global minima of the distance between the contour points and the center area point. Such global minima is representative of a break in the contour which can be an indication of alga fission. First, we detect the local minima. Among these, the global ones to a section are searched. Then, we eliminate those that are not representative of a break in the contour.

All along this step, we have been kept in memory the size of the different objects. The application of the previous threshold will be more severe for a great size than for a small one.

Alga modelling : If we use a simple right cut applied on the break point level, we will lose a part of shape information. The classification will be more difficult, as the feature extraction. We prefer try to model algae from the partial contour information.

As we know the break point coordinates, we have access to the coordinates of all points composing each alga sections. The algae are quite circular, so we decide to use an quadratic equation to modelize each bit of contour. We choose the Least Mean Square method to calculate the unknown quadratic cœfficients.

Are eliminated as an alga all the point sections giving another equation solution rather than an ellipse. This case is rare enough and in general due to a wrong number (superior one) of contour breaks.



5 - a 5 - b

Figure 5 : After the additional filtering, we obtain : (a) a real global contour image. The modelling is performing to achieve modelled algae (b) if necessary.

Are also eliminated the ellipses which have an area too great or too small in relation to the aggregate area. This case is rare enough too. The elimination of one modelled alga in an aggregate is not a problem because the decision on the kind of algae in the determination step will be on all modelled algae composing each aggregate.

Modelled alga feature extraction : The parameters as above can be extracted in the same way to have comparative results. But as the algae in cellular division are smaller than the single ones, applying the same classification scheme seems to be difficult. The separation is better with the use of alga models, but not enough separative. We have a certain translation on the size axis and smaller one on the elongation axis. So we must create other classes for modelled algae.

The last parameter, applicable only on aggregates, can be extracted. This is the fission angular separation between the algae inside the aggregates. If the alga strain is quite circular, this separation will be perpetuated perpendicularly to the major inertia axis of the modelled algae. If the modelled alga is in rocket form, less obvious difference with the previous modelled alga, the separation will be made along the major inertia axis of the aggregate. The angular distance between the major inertia axis of the modelled algae and the major inertia axis of the aggregate is minimal.

This property is known to be difficult to see in light microscopy by a human. In fact, on light microscopy binary images, it seems to be relatively easy to extract this feature by this way.

IV. CLASSIFICATION AND IDENTIFICATION

The use of the principal component analysis permits us to select a discriminant feature set. A feature vector, attached to each alga, is formed. As we have a small sample number of each genus, ranging from 30 to 60 individuals per class, we turn towards a classical binary decision tree.

The general process is as follow. We select a few set of algae. This composed our learning set. With this set, we construct the general tree architecture. Then we test the architecture validity with the rest of algae.

The first method used is a classical one. At each step of the tree construction, we employ the alga histogram distribution of each vector feature field to determine which is the most discriminant. The most separative field is selected to be the decision parameter. The detected threshold values are the new decision threshold values. A process permitting to recover some bad classification is integrated between two steps. We test the final tree architecture with the rest of alga feature vectors. The results of this first method is represented in figure 6.

%	A	B	C	D	E
A	87,5	10	13,2	-	-
B	-	90	10,5	-	-
C	-	-	76,3	-	-
D	-	-	-	86,9	9,4
E	12,5	-	-	13,1	90,6

Figure 6 : Table of the first method results. This table shows the well identification rates as well as the miss-classification rates. The A to E letters represent alga genera.

The rates of well identification are high enough. The rates of miss-classified algae are relatively important. Some algae are confused. These difficulties could be anticipated. The apical view of some rocket shape algae can be very close to round algae of the same size. Another similar problem can be detected with bigger algae. Some Gloeothece view seem alike a Microcystis view. This problem is due to the 2D projection. Perhaps some textural feature parameters could help to remove the ambiguity.

The second tested method is an adaptive one. The learning process is the same as the previous one in the general tree architecture. On the other hand, the threshold values are not fixed. They are calculated from the entire alga set. This new method permits to adapt the threshold to have a general classification process. The classification results are shown on figure 7. The results are similar to the precedent ones. However the good identification rates seem to be more regular.

Genus / Classif.	A	B	C	D	E
Learning	100%	100%	100%	100%	100%
Algae	81,3%	80%	92%	89,8%	84,9%

Figure 7 : Table of the second method results. This table shows the well identification rates . The A to E letters represent alga genera.

As a conclusion, the well identification rates are good enough for each of the two classification methods. Some problems exist, but they could be anticipated. The algae situated in the neighborhood of each alga to be classified could be an interesting information . This information could be used to decrease the miss-classified alga rates.

V. CONCLUSION

We have first presented a general method of object location on light microscope images. We tried to use all the information present in the images without modifying them. Then, we select the contour corresponding to the external object contours. So we are able to extract object shape features. We obtain parameter vectors which can be employed in the classification process. The general sequence of the described process can be summarized as in the figure 8.

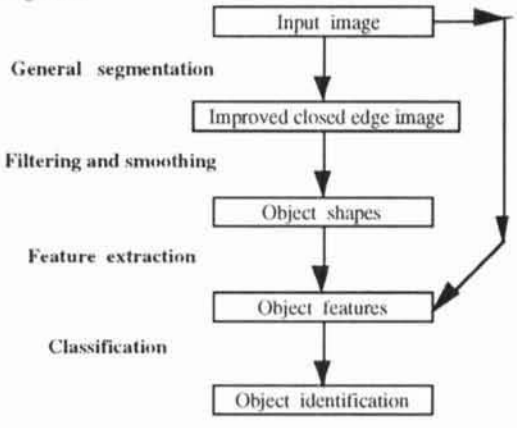


Figure 8 : General sequence of the different steps used.

The general segmentation and additional filtering permits us to recover about more than 90% of the external alga contours on images. Although the set of algae is not so important, we reach more than 80% good alga identification rate. These results are good comparing to the poor quality of the input images.

The bad results can be explained in part by the use of only 2D shape features. Additional grey level tone features can be employed to increase the good identification results, as well as to detect and to eliminate non confocal algae. In the same way, the use of the global neighborhood information could even perform better results.

REFERENCES

1. Rippka R., Deruelles J., "Generic Assignements, Strain Histories and Properties of Pure Cultures of Cyanobacteria", Journal of General Microbiology, 111, pp. 1-61, 1979.

2. Roussel M., Fontaine D., Tu X.W., "Biological object location on light microscopy images", SPIE Biomedical Image Processing and Three-Dimensional Microscopy, Vol. 1660, pp. 388-396, 1992.

3. Canny J., "A Computationnal Approach to Edge Detection", IEEE PAMI, Vol. 8, no. 6, pp. 679-697, Nov. 1986.

4. Deriche R., "Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector", International Journal of Computer Vision, pp. 167-187, Kluwer Academic Publishers, Boston, 1987.

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

6. Freeman H.,"On the Encoding of Arbitrary Geometric Configurations", I.R.E. Trans. Electron. Comput., EC-10, n° 2, pp. 260-268, 1961.

7. Kitchin P.W.,"Processing of Binary Images", Robot Vision, edt. A. Pugh, IFS Ltd., U.K., Springer_Verlag, pp. 21-42, 1983.