**WASTE WATER ALGAE DETECTION USING IMAGE PROCESSING**

**I.INTRODUCTION**

Microalgae have been considered as the most promising candidate for future generation. Recently it emerged as light owing consumer product in various industry. Profitable products like biomass for food and animal feed, valuable yields like triglycerides can be produced from algae. Likewise algae have also been used for purification of heavy water by accumulating heavy metals, organic and inorganic substances in their body. In addition detected algae are harvested in treatment ponds for large scale production and are used to control water degradation. It has varying lipid content and this property makes it useful for bio-fuel production. So for large scale production of this useful organism industrially identification of the organism is very crucial as the manual identification process is time consuming and expensive.

So, Automated Computer Based identification system of microorganisms from their image will reduce the burden of

manual process which involves more man power, cost and is also time consuming.

This automated system is advantageous because of its learning capability from a given dataset (images).

**II.METHODS**

MATLAB 7.9.0 (R2009b) is used to develop an automated waste water algae detection and classification system.

The simulation platform has the ability to give a computational environment to manual methods. It’s a high-level programming language aided by different functions and toolboxes. An image processing tool box is used in this paper. The methods used to develop this automated recognition prototype include image pre-processing, segmentation, feature extraction, clustering using the Fuzzy C-means method and optimization techniques. The system architecture is given in Figure1.

1. ***Image preprocessing***

Images collected through microscopic photograph are mostly affected by noise. They may also have low contrast and unsharpened edges. So, we need to enhance the image and make it noise free for further analysis.

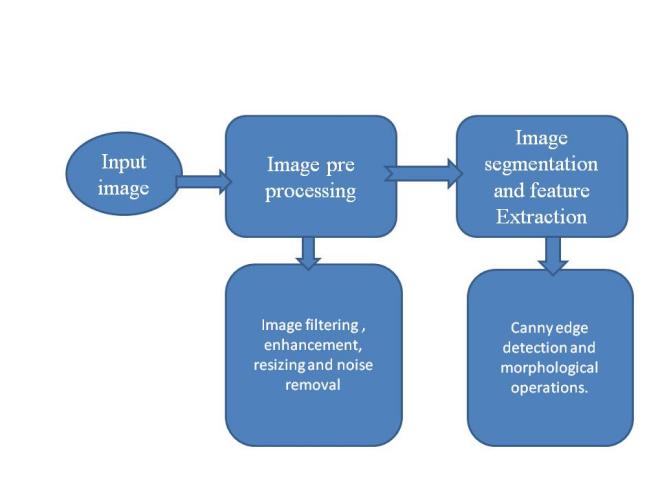


Figure 1.

The steps involved in image pre-processing are:-

1. Images are captured and uploaded to the system using Graphical User Interface (GUI).
2. Image was converted from RGB to gray scale.
3. Image enhancement and sharpening was performed to adjust the image brightness, contrast and sharpen the images.
4. As images were converted from rgb to gray scale we need to maintain the intensity of pixels. So, Histogram equalization was performed and pixel intensities are mapped to a graphical representation.
5. Conversion of gray scale image to binary image.
6. ***Image segmentation***

It is a method of partitioning image into different segments (clusters). The main objective is to modify the picture representation that is more analytical. It is used to detect boundaries or isolate a particular part. Here, a label is assigned to each pixel and cluster groups are performed by aggregating similar pixels.

Previously obtained pre-processed binary image is input for segmentation process.

Canny edge detection algorithm is used for boundary extraction. It calculates the discontinuity in image intensity values and hence boundary is obtained.

Input image is divided into multiple images by object detection called sub images. Each sub image refers to one object only. Enclosures of 60 pixels in length and a minimum of 40-50 pixels perpendicular to it from the original binary image forms a new sub image. The resultant sub image undergoes subnet masking to obtain same region of original image by referring the pixel intensity value and finally gets stored in a database.

***1) Feature Extraction***

It transforms or converts the image to a set of parameters that represent features of algae. The parameters included in this extraction process are shape, area, minor and major axes, perimeter, and intensity spectrum with Principal Component Analysis.

The steps involved are: -

1-Median filter was applied to smoothen the image and make it noise free.

2-Local Gradient (1) and Edge Detection (2) were computed at each point.

L(x , y )= [lx2 + ly2] (1)

Where lx and ly are first-order derivatives of the intensity of pixels.

E(x, y) = tan-1(lx/ly) (2)

3- Minimum convergence in the gradient magnitude yields a thin line which is a ridge formed by the combination of coordinates of the boundary pixels.

4-Some morphological image analysis like erosion, dilation or masking is performed according to the image requirement.

***2) Shape Determination***

There are algae of different shapes, some may be circular or irregular shapes or spirals. To categorize them based on their shape, we need to label common shapes like 0 for circular, 1 for spiral and -1 for irregular shapes. Hence function is used to distinguish the shape. Thus, the function improves the accuracy rate and makes image recognition efficient and easy.

First, we calculate the angle of alignment of the organism’s image with the horizontal axis. Then the algae are rotated according to the experiment requirement. The angle of inclination is calculated by obtaining the longest path between each two algae points on the algae boundary. The two points P1(X1, Y1), and P2(X2, Y2) are two points on edge of the algae. The origin is P(0,0).

Angle of Inclination (Ai)= tan-1(d1-d2/1+(d1-d2))

Where d1=(Y2-Y1)/(X2-X1) and d2=(Y1-Y0)/(X1-X0).

***3) Perimeter Extraction***

The perimeter of the image is calculated with the help of intensity pixels of the image is calculated with the help of intensity pixels of the image boundary. The distance between adjoining intensity pixels on the edge of the image is calculated.

The perimeter is given by

I Pi=i

Where di = distance between adjoining intensity pixels in the edge of the image.

***4) Area***

The area is calculated from the binary image by finding the summation of all the white pixels within the edge of the image.

***5)Major axes and minor axes***

Major axes are represented by joining the end pixels of an image and the axes perpendicular to it are the minor axis. In other words, the minor axis of the algae image represents the width and the major axis represents the length.

1. ***K-Means Clustering Algorithm***

In order to understand k-means clustering, we have to first know about clustering. A cluster is a group of data items that have been combined because they have similar characteristics. K-means clustering basically helps us to group similar data and understand the pattern of the clusters so formed. When we have a huge data set, we form clusters of data in order to simplify it. Every data point that we have, is allocated to a cluster. Thus, this reduces the in-cluster of the sum of squares. Now, these clusters so formed have centroids, and the data point is allocated to the nearest cluster while keeping the centroid as small as possible.

So, to sum up, the ‘K’ in k-means is for the number of clusters to be formed and ‘means’ refers to the averaged-out data or finding the centroid of the clusters so formed.

Now, in k-means clustering, we first take a set of observations, say, (x1, x2, x3…., xn).

Here, we have considered d-dimensional real vectors with n observations such that k≤n.

Let S= {S1, S2,S3,…SX}

argsmin ||k-μi||2 = argsmin|Si|var Si.

Where μi is the mean of points in Si.

Now,

argsmin 

The equivalence can be deduced from the identity:



Coming to the algorithm of k-means, mathematically speaking, it means partitioning the observations according to the Voronoi diagram generated by the means:



Where each xp is assigned to exactly one s(t), even if it could be assigned to two or more.

Recalculate means (centroids) for observations assigned to each cluster

mi(t+1)= 

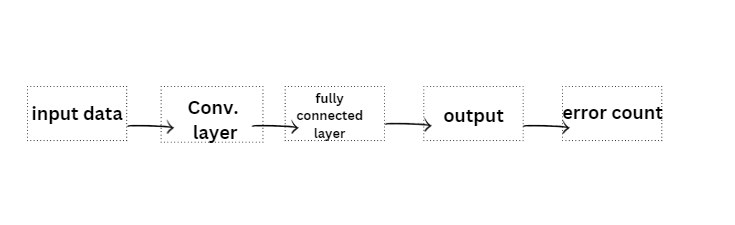
GLOBAL OPTIMIZATION AND METAHEURISTICS

The classical k-means algorithm and its variations are known to only provide coverage to local minima of the minimum-sum-of-squares. The clustering problem is defined as:



1. ***CNN***

Convolution Neural Network is a part of Deep Learning which is a part of Machine Learning. It has a variety of applications but one of the main applications of CNN is image recognition and pixel data processing. There are plenty of image recognition and pixel data processing neural networks, but CNN works the best. This is because CNN is kind of a network architecture. The basic working of CNN is given as follows:



2D CNN RESULT MATRIX





3D CNN RESULT MATRIX



DIMENSIONS OF OUTPUT MATRIX



CONVOLUTION LAYER





Were,

intermediate value

containing filters or weight

bias value

activation function

LAYER BACKPROPORTION



The filter is denoted by , and  is a scalar that belongs to a partial derivative



CALCULATING THE DERIVATIVE OF THE ERROR

Operation applied in one layer

For layer 1:

X1-----> y1----->a1

Input- x1=a1-1

Process- y1=x1\*w1

Activation- x1+1=a1

For input: x

Dimension: 

A number of input planes: 

Size of each plane: 

The convolution filter is denoted by: 

Dimension: 

A number of filters: 

For output:  There are j planes of size 

amount of zero padding

Stride length, 1 in our case

 input channels, 

output channels, 

Layer, only 1 layer in our case

Identify activation function 

ERROR FUNCTION

e(y)=all pixels in y j(yi-ŷj)2 , 

z(yj-ŷj), (n-w+1) (n-w+1)

p = 0, S = 1

2(yj-ŷ)\*flip(xi)

1. **RESULTS**

The developed system uses image segmentation and enhancement and has been used in several images. It is applied to images of 35 organisms and is validated to be efficient in its performance. It achieved around 92% efficiency by successfully identifying the organisms based on their texture, shape like circular, spiral etc. It also distinguishes the algae from other unwanted objects. 45-48%. This work has concentrated on the implementation of robust algorithms which can be tested on representative samples of data.

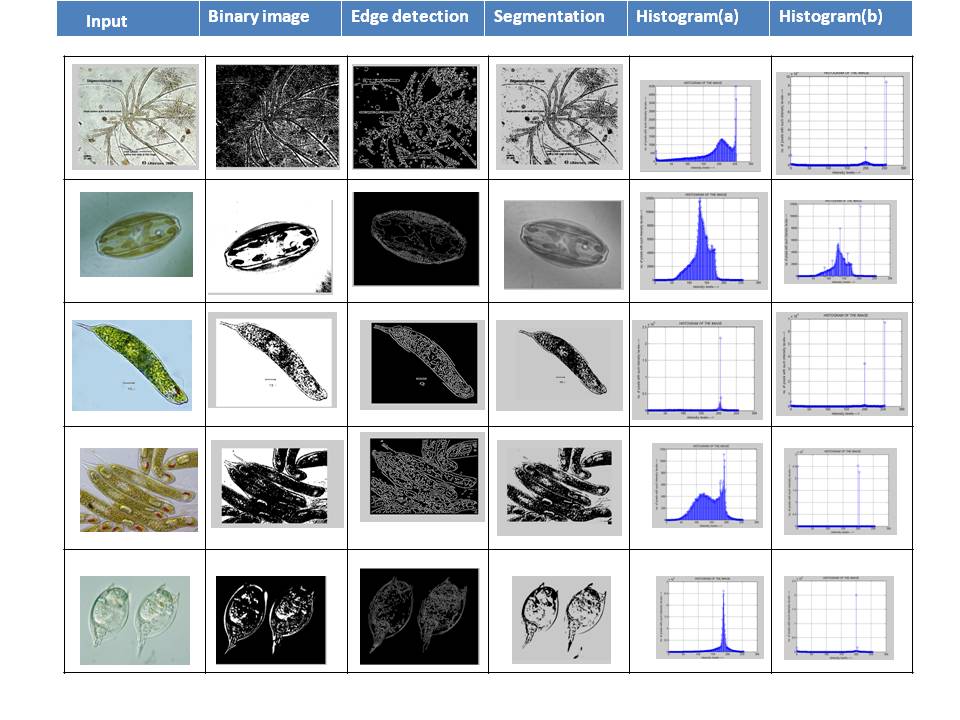


Figure 2. Image pre-processing and segmentation process

**IV.DISCUSSION**

Our main objective was to develop an automated computerized system to recognize algae. Here image is used to detect algae by segmenting the image into sub images and then analysing the sub images. Analysis is not beneficial on sub images because for the complete image the input data set will be very large and obtained result is not in context with the required output. So, image is divided into sub images and aligned with their original position. This work is an illustration of automated process to reduce the number of manual candidateship. The colour image is converted to binary and then segmented using Fuzzy C Means Clustering on the pixels. Lastly Histogram Equalization is performed to show the graphical representation of processed images.

1. **CONCLUSION**

We have presented a computer aided automated algae detection system for waste water management using image processing and clustering. The proposed method detects confocal images with maximum accuracy with respect to their histological standard. Processing images can utilize the time history information obtained from the image sequence, but then utilizes feature extraction, edge detection techniques, filtering methods, and segmentation techniques.

1. **REFERENCES**