Automatic In Situ Identification of Plankton

M. B. Blaschko, G. Holness, M. A. Mattar, D. Lisin, P. E. Utgoff, A. R. Hanson, H. Schultz, E. M. Riseman
Department of Computer Science, University of Massachusetts Amherst
M. E. Sieracki, W. M. Balch, and B. Tupper
Bigelow Laboratory for Ocean Sciences

Department of Computer Science
University of Massachusetts
140 Governors Drive
Amherst, MA 01003

Bigelow Laboratory for Ocean Sciences
PO Box 475
West Boothbay Harbor, ME 04575

http://vis-www.cs.umass.edu/projects/bigelow/

Abstract

Earth's oceans are a soup of living micro-organisms known as plankton. As the foundation of the food chain for marine life, plankton are also an integral component of the global carbon cycle which regulates the planet's temperature. We present here a technique for automatic identification of plankton using a variety of features and classification methods including decision trees, ridge regression, k-nearest neighbor, support vector machines, and ensembles. The images were obtained in situ by an instrument known as the Flow Cytometer And Microscope (FlowCAM), that detects particles from a stream of water siphoned directly from the ocean. The images are of necessity of limited resolution, making their identification a rather difficult challenge. We expect that upon completion, our system will become a useful tool for marine biologists to assess the health of the world's oceans.

1 FlowCAM

Figure 1 Portable FlowCAM system

An instrument for monitoring the abundance of phytoplankton and small zooplankton, the Flow Cytometer And Microscope (FlowCAM) [2], detects and takes images of micro-organisms from a stream of water siphoned directly from the ocean. It has a rudimentary image segmentation capability, used to crop out individual organisms. The instrument is used by marine biologists to estimate the population sizes of different plankton species. In particular, scientists are interested in potentially harmful organisms. At present scientists must classify images by hand, a tedious and time consuming process.

2 System Layout

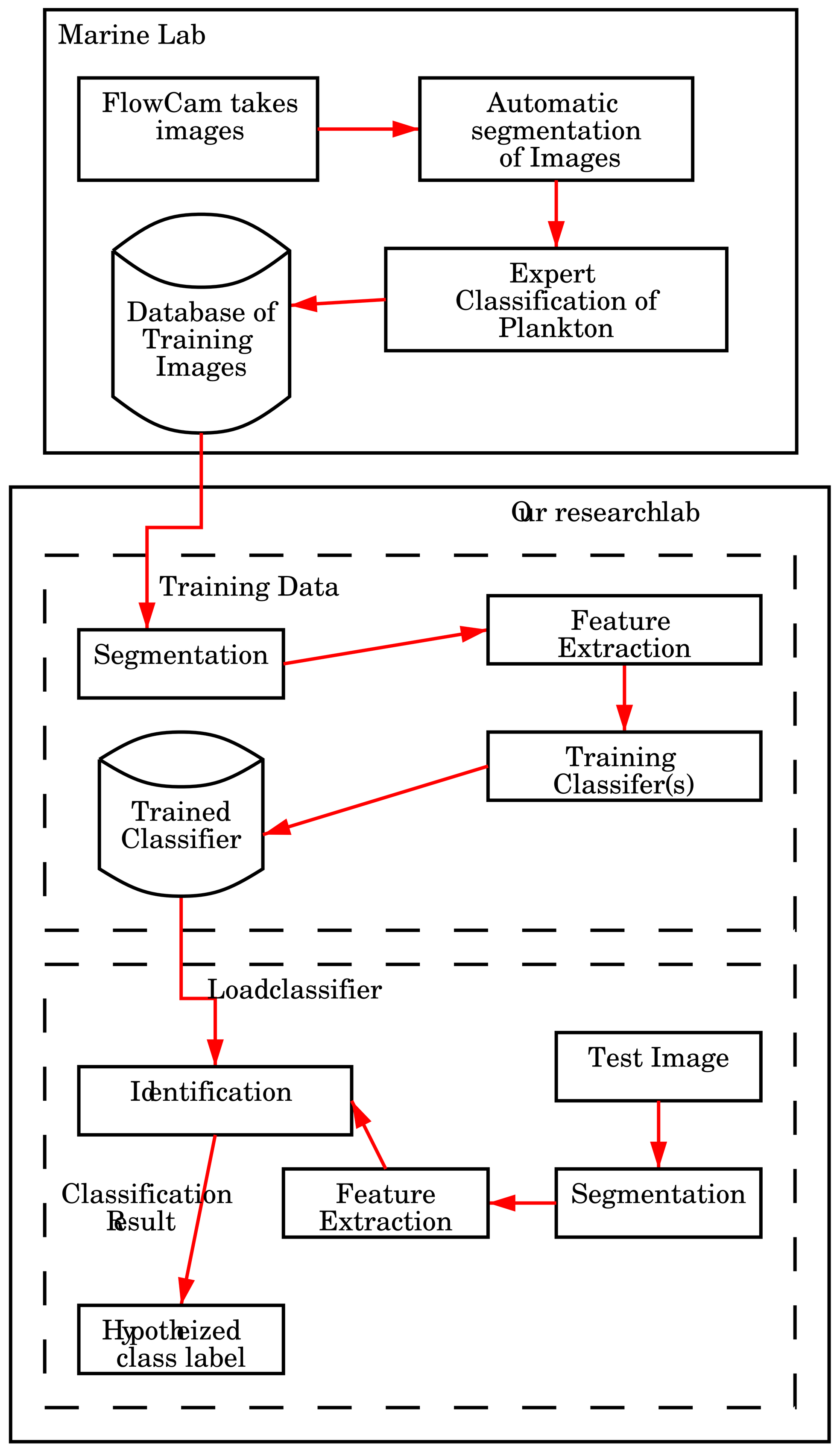

Figure 2 System layout diagram

Our data set consists of FlowCAM generated images that have been expertly classified by marine scientists. The classifications correspond to broad groupings that reflect the general distinctions useful to marine scientists, and that often (but not always) correspond to biological and visual similarity. After receiving plankton images from the Marine Lab, we perform segmentation techniques. Then we extract the features from these segmented images and train classifiers. At this point, the trained classifier is saved locally and ready for test images. A test image follows the same processing steps, its features are sent to the trained classifier, and a class hypothesis is returned.

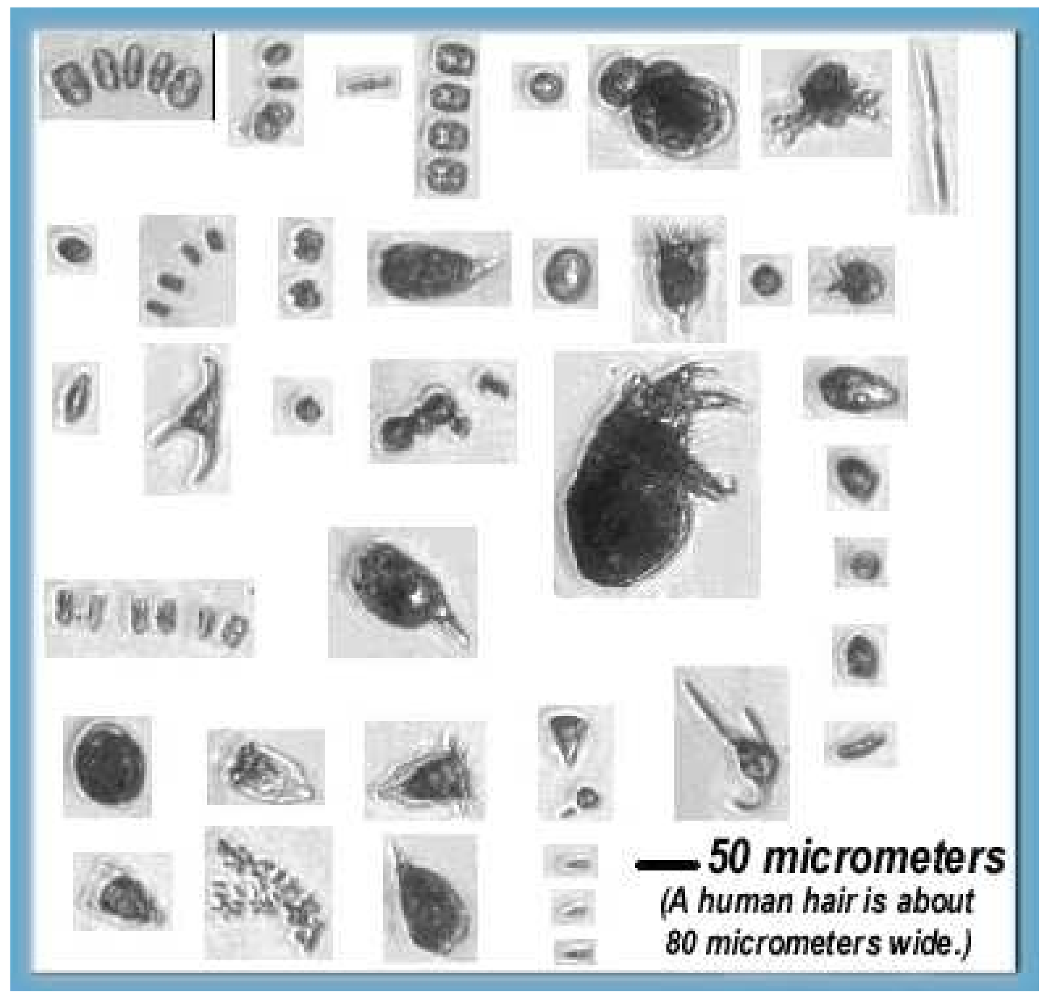

Figure 3 Example FlowCAM imaged aquatic particles

3 Segmentation

Active contours or deformable models, also known as snakes [1], are energy minimizing splines that can move under the influence of a potential field computed over the intensity surface of an image. The potential field exerts a force such that the snake comes to rest at the boundaries of desired features.

We have also found that a simple global bimodal segmentation is effective in many cases for separating the plankton from the background, which tends to be significantly brighter than the object.

Both segmentation methods are used in this paper as a preprocessing step for extracting different features. The intensity-based segmentation is better at capturing fine details, while the snake-based segmentation is guaranteed to produce a smooth closed contour.

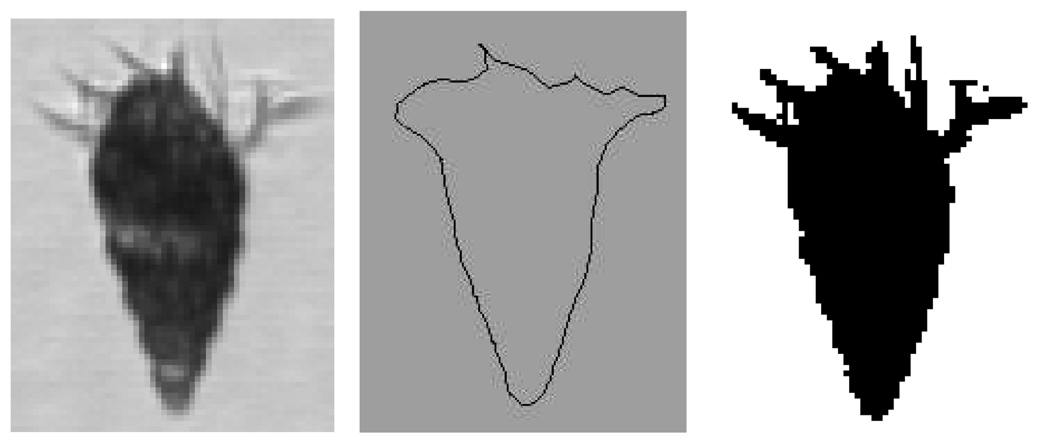

Figure 4 A sample image (left), its snake-based segmentation, and its intensity-based segmentation (right)

4 Feature Extraction

Features were grouped into five types: simple shape, moments, contour representations, differential and texture features. This grouping reflects different ways, in which the image is represented.

Category	Feature Set
Simple Shape	Simple Shape (SS) [9]
Moments	Moment Invariants (MI) [7]
	Zernike / P-Zernike (Zke / P-Zke) [32]
Contour	R-θ Hough [128]
	Fourier [10]
Differential	Shape Index (SI) [120]
Texture	Size, Mean, Variance (SMV) [3]
	Co-occurrence (CO) [140]
	Local Binary Patterns (LBP) [54]

The number in the square brackets reflects the number of features used from each feature set.

5 Classification

As shown in the following table, a combination of different feature types results in approximately 70% accuracy with Support Vector Machine classification.

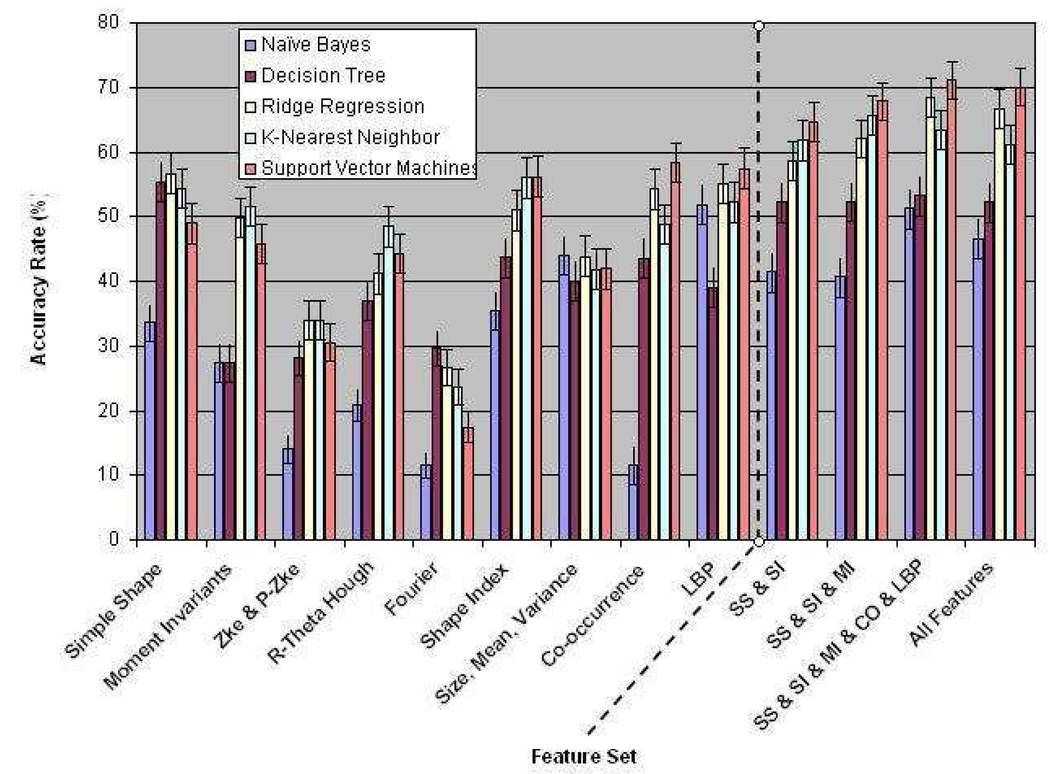

Figure 5 Classification results for various feature selections

References

[1] M. Kass, A. Witkin, and D. Terzopoulos. Active contour models. *International Journal Computer Vision*, 1(4), 1988.

[2] C. Sieracki, M. Sieracki, and C. Yentsch. An imaging-in-flow system for automated analysis of marine microplankton. *Mar. Ecol. Progr. Ser.*, 168:285–296, 1998.

This work is funded by the National Science Foundation under grant number ATM-0325167.