CLASSIFICATION OF CHALLENGING MARINE IMAGERY

Piyanuch Silapachote, Frank R. Stolle, Allen R. Hanson
Department of Computer Science, University of Massachusetts Amherst, Amherst, MA 01003, U.S.A.

Cynthia H. Pilskaln
School of Marine Science and Technology, University of Massachusetts Darthmouth, North Darthmouth, MA 02747, U.S.A.

Keywords: Marine science application, Biologically inspired vision system, Image segmentation.

Abstract: Covering over 70% of the Earth's surface and containing over 95% of the planet's water, the aquatic ecosystem has a great influence on many environmental functions. An indicator of the health of a marine habitat is its populations, estimated by taking underwater images and labeling various species. Designing an automated algorithm for this task is quite a challenge. Image quality tends to be low due to the dynamics of the water body. The diversity of shapes and motions among living plankton and non-living detritus are remarkable. We have applied two very different techniques from computer vision to the automatic labeling of tiny planktonic organisms. One is a common approach involving segmentation and calculations of statistical features. The other is inspired by the sophisticated visual processing in primates. Both achieved competitively high accuracies, comparable to general agreement among expert marine scientists. We found that a relatively simple biologically motivated system can be as effective as a more complicated classical schema in this domain.

1 INTRODUCTION

Self-regulating marine ecosystems are home to many primary producers. There is a need to compile information on the composition and size distribution of marine phyto- and zooplankton as well as the organic-rich, aggregated detrital material resulting from life activities and death throughout the oceans (Benfield et al., 2007). Plankton mediate the flow of carbon from the atmosphere to the oceans and strongly influence aquatic bio-geochemical cycling and nutrient budgets. The response of planktonic ecosystems to global climate change is being documented in decadal changes in production levels, plankton community composition, and carbon flux through the ocean (Hays et al., 2005; Honjo et al., 2008).

To formulate predictive models of future marine communities and their influence on oceanic sequestration of atmospheric CO_2, oceanographers require a large amount of spatially and temporally variable data on marine food webs and the cycling of carbon and nitrogen through them. Automated analysis of plankton and particle imagery collected worldwide would provide the necessary data stream.

Identification of marine populations is a fundamentally challenging research. Living organisms and non-living particles come in significantly diversified shapes and may undergo highly articulated motions. A number of organisms are microscopic. Particles are continuously drifting. Added to the problem is getting high quality images when both sensors and objects may be moving and the medium is an unstable flow of water current. Manually labeling these complex scenes is not only very laborious and time consuming, but also very difficult even for human experts. In one study expert marine scientists agreed on labels only 75-80% of the time (Culverhouse et al., 2003).

We believe that marine biologists can benefit from an application of recent Computer Vision techniques. Meanwhile, marine imaging provides a new data source with distinctive characteristics that may not be represented in other data sets. As discussed in (Ponce et al., 2006), databases currently available and commonly used for evaluating multi-class categorization, e.g. the Caltech set, lack irregularities within a single category. By contrast, taking an average over all images of each class in our marine data (examples shown in Figure 1) reveals high intra-class variations. Objects exhibit no conformity in size, shape, pose, or orientation. They do not share a uniform location. Vi-

Silapachote P., R. Stolle F., R. Hanson A. and Pilskaln C. (2010).
CLASSIFICATION OF CHALLENGING MARINE IMAGERY.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 401-406
DOI: 10.5220/0002835304010406

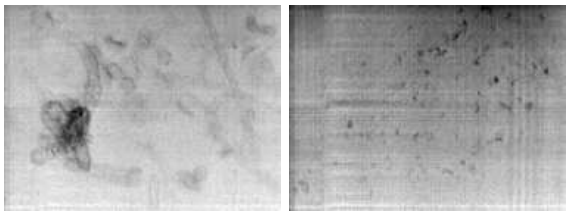

Figure 1: Average images of chain colonial radiolarians (left) and marine snow from our data set. See Figure 4 for individual samples. Images throughout this paper are enhanced for display by inverting and/or adjusting brightness.

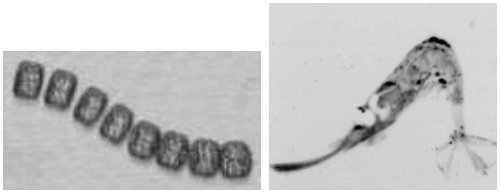

Figure 2: Images from FlowCAM (left) and VPR (right).

sually recognizing marine species from their average images is extremely difficult, if not impossible.

We applied two very different techniques to automatically label taxonomic categories of challenging marine images. The first, referred to as the *traditional approach*, is common in the machine vision community. Processing involves image segmentation, compact feature vector representation, followed by classification. Through many years of research, segmenting an image has remained a non-trivial problem. Solutions are usually not well-defined and most often are highly dependent on specific data sets and their applications. Given the unique characteristics of our plankton images, we propose a novel segmentation technique that is capable of effectively handling very complex underwater scenes of microscopic organisms.

The second approach, referred to as the *biological approach*, is based on (Serre et al., 2005) and is inspired by the success of sophisticated visual systems in primates. With only 2 main computations, correlation mimicking simple cells and maximization mimicking complex cells, the model is relatively simple. Features are constructed from randomly selected image regions. Neither segmentation nor focus of attention is necessary. Unlike the traditional approach, this domain independent trait makes the biological system effectively generalizable. This benefit comes with a high computational cost, which is far from being real-time. Investigation of the model complexities led to modifications that resulted in a substantial computational reduction and an improvement in classification.

2 RELATED WORKS

2.1 Plankton Classification

Two of the most related works are (Blaschko et al., 2005) and (Lisin, 2006). Blaschko et al. employed a variety of existing features and classifiers. They identified plankton in images obtained using a Flow Cytometer And Microscope, FlowCAM (Sieracki et al., 1998), on which their accuracy was 72.61%. Lisin studied a kernel density estimation on local bags of features and proposed a methodology combing multiple global and local features. His algorithm improved the results obtained by Blaschko et al. on the same FlowCAM data, achieving 74.84%. He also applied his technique to images obtained using a VPR (Video Plankton Recorder), reporting 71.90% accuracy.

Here, we look at a difficult set of video sequences of several species of tiny underwater organisms (refer to Section 5 for a complete detail). Our images were obtained using an older VPR that exhibited significantly greater noise than the VPR used by Lisin.

There are significant differences among the three marine image sets: FlowCAM, Lisin's, and our VPR. Samples of actual images used by Blaschko et al. and Lisin are shown in Figure 2. Images are mostly well conditioned. An organism is clearly presented as foreground, though some are partially cropped, and fine details are apparent. A relatively high contrast simplifies background separation and segmentation.

By comparison, as evidenced in Figure 4, our images are particularly challenging due to variations within and between classes, as well as low image quality and limited resolution. The data exhibit a low signal-to-noise ratio and significant correlated noise. Detecting and classifying objects are rather difficult in these images even by humans. Organisms are mostly localized and typically occupy only small regions. Many are not captured in their entirety. A number of organisms are translucent resulting in regions that are combinations of foreground and background. Some are distributed, appearing as separate particles, often in extremely diversified shapes.

2.2 Biologically Inspired Vision System

Primate effortlessly perceive and efficiently use visual cues to extract reliable information from less than pristine data. How this fascinating process is executed has long been an intriguing question in multiple disciplines, most notably, neuroscience, physiology, and psychology. Early pioneers, (Hubel and Wiesel, 1959) discovered simple and complex cells and described them as edge detectors. Spatially larger, com-

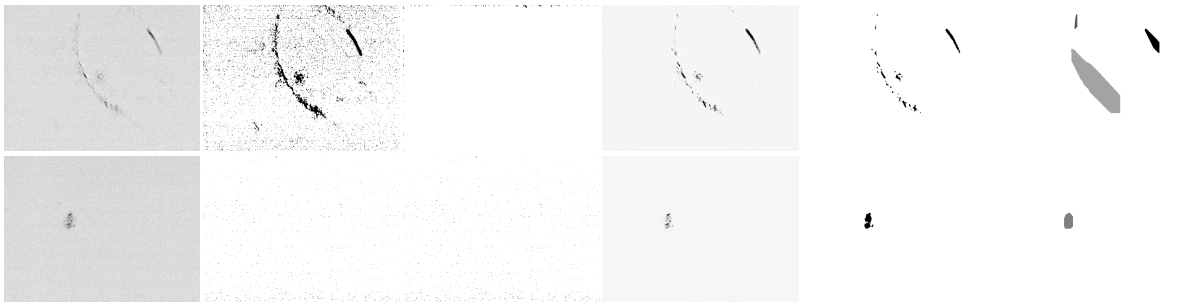

Figure 3: Our segmentation applied to two examples of marine images, one on each row, side by side with results of the technique used by (Lisin, 2006). Shown from left to right are the original image, the intermediate and the final results of Lisin's approach, and results of our three-step approach, after noise removal, candidate segmentation, and grouping. While Lisin's technique fails on these images, our approach successfully segments both organisms from their background.

plex cells are less sensitive to transitional changes. Later, (De Valois and De Valois, 1988) suggested a profile of receptive fields of simple and complex cells, a structure that resembles Gabor functions.

A model of a biologically feasible vision system was designed by (Riesenhuber and Poggio, 1999). They emphasized the two fundamentals: specificity and invariance. Their hierarchical model, HMAX, consisted of linear S layers performing template matching and nonlinear C layers performing a MAX-like pooling. They introduced and detailed how the MAX operation could be a key mechanism for invariance when building complex cells from simple cells.

Serre et al. extended HMAX beyond the cortical neurons (Serre et al., 2005). On commonly used data sets, this enhanced system, referred to as the Standard Model (more detail in Section 4), performed competitively with respect to other state-of-the-art techniques.

3 TRADITIONAL APPROACH

3.1 Image Segmentation

A customized segmentation is developed to overcome the challenges associated with the data. After significant exploration of alternative methods, the best results (Figure 3) are obtained using a 3-step approach: noise removal, candidate segmentation, and grouping.

Fractional spline wavelet filtering (Blu and Unser, 2003) is employed to clean up noise. A wavelet transform of an image is computed, then thresholded on the coefficients. A single threshold is empirically determined for all images in the set. Finally, the image is reconstructed from its thresholded wavelet transform. A wavelet degree of 3 is used with a threshold of 50.

Noise removal is followed by a segmentation algorithm originally developed to count nuclei in digital microscopic images (Byun et al., 2006). This candidate segmentation consists of 4 steps: median filtering, thresholding, watershed filtering, and size filtering. Median filtering with a 5x5 window is performed to remove more of the remaining noise. Automatic thresholding attempts to find an optimal threshold by minimizing intra-class variance with respect to particles and background. The resulting binary image is processed with watershed filtering and a morphological opening operator to remove spurious small particles, then a size filter is applied to remove very large ($>$5000 pixels) and very small ($<$5 pixels) particles.

The last step groups all candidate particles within a certain radius, increasing the particle area and attempting to restore coverage for areas that were removed in previous steps. This improves segmentation as translucent parts of many organisms are initially removed. However, a sufficient number of widely distributed parts still remain to indicate the outline of the particle. A grouping distance of 50 pixels is used, based on image resolution and expected object size.

3.2 Feature Extraction and Classification

Several types of features are extracted after segmentation: simple shape descriptors, moments, contour representations, texture features, and shape indices. This set is similar to the set used in (Blaschko et al., 2005).

Simple shapes, computed from each of the segmented particles, include area, perimeter, compactness, ratio of eigenvalues, eccentricity, rectangularity, and convexity. Grayscale moments computed are mean, variance, skewness, and kurtosis. In addition, moment invariants as defined by (Hu, 1962) over binary as well as original grayscale images are used.

A set of texture descriptors is derived from co-occurrence matrices covering the segmented particles. A co-occurrence matrix is a two-dimensional histogram that tracks the number of occurrences of pairs

of graylevels at each horizontal and vertical displacements. From these matrices, energy, inertia, entropy, and homogeneity (Haralick, 1979) were calculated as features. Other texture descriptors were local binary patterns (Ojala et al., 2002), derived by calculating a circle of points around a center pixel and assigning binary weights to the signs of the differences between these pixels and the center pixel. These patterns are calculated for each pixel in the segmented particle and are accumulated into a histogram of counts. Normalizing by object size yields the features used.

Shape indices are effective global differential image descriptors for object recognition (Ravela, 2003). They are functions of the isophote and flowline curvatures of the image intensity surface, which are computed using Gaussian derivative filters. Shape indices are computed for every pixel in the image, including background. Histograms of individual shape indices are calculated and used as features.

Classification is done using a Support Vector Machine (SVM) with a third order polynomial kernel. The results are based on 10-fold cross validation.

4 BIOLOGICAL APPROACH

The Standard Model (Serre et al., 2005), a framework for constructing a set of image features, is designed to be consistent with the structure commonly agreed to exist in the immediate visual processing of primates.

Primates have a remarkable ability to recognize an object that has undergone various transformations, such as affine and perspective projections, or extreme changes in lighting conditions, and to distinguish similar objects belonging to different categories. Underlying this effectiveness is a balance between selectivity and invariance. The Standard Model mimics these mechanisms by interleaving simple S with complex C layers in a hierarchical feedforward architecture. Simple cells perform selectivity, responding strongly only to patterns they are tuned to. Complex cells are tolerant to changes in scale and translation. They nonlinearly combine activations and inhibitions from multiple similarly tuned simple cells.

This structure supports a gradual increase in complexity and invariance of neurons as well as the size of their receptive fields along the visual pathway. While progressing from primitive tokens to defined shapes and forms, MAX pooling covers larger image regions.

4.1 The Four-layer Standard Model

The first layer, S1, resembles simple cells in V1 that process spatial frequency information. S1 is modeled by Gabor filters, extracting edge-like features at 16 scales and 4 orientations. The second layer, C1, simulates complex cells. Neighboring scales and neighboring pixels of S1 maps are combined using a MAX operator. Subsampling these responses creates C1 maps.

During training, image patches are randomly extracted from a number of randomly selected C1 maps. Given an unknown input image, locating an area that best matches with each of these patches is a process in S2 and C2. A patch is convolved with C1 maps of the input, generating S2 maps. Maximizing across all scales and positions results in a C2 value. Intuitively, a feature value indicates how similar a region in an image is to a particular patch. This procedure is repeated for every patch, thus constructing for each input a feature vector of length equal to the number of training patches extracted. Feature vectors are then classified using a supervised one-against-one SVM.

4.2 Modification to the Standard Model

One drawback of the Standard Model is its high computational cost precluding real-time operation. Additionally, with the amount of randomness involved, an experiment needs to be run repeatedly in order to obtain good statistical estimates. Empirically tracing the computations, the bottleneck is in correlating training patches with images. We apply two approaches to reduce the number of patches and thus overall computational cost without sacrificing performance.

Overlaps among Image Patches. Given the size of an image and a patch, there is a limited number of patches, n, that can be extracted so no patches overlap, for example, n is 12 when extracting 32×32 patches from a 128×96 image. Partial overlaps capture different views or different parts of an image and how they are fused together. Excessive overlaps, on the other hand, lead to unnecessary duplicates and increase dependency among patches. We prevent an image from being overly represented with overlapping regions by extracting a maximum of n patches per image. Instead of exclusively selecting non-overlapping patches, we maintain random selection allowing partial overlaps.

Information Contained in an Image Patch. Generally, a classifier needs to be trained on both foreground and background information. A few background patches are necessary, but not so many that they significantly increase computations and introduce bias. Our marine data is sensitive to the latter problem due to uniform underwater background and organisms occupying only a small area on an image. A large number of extracted regions, especially smaller ones, are exclusively background.

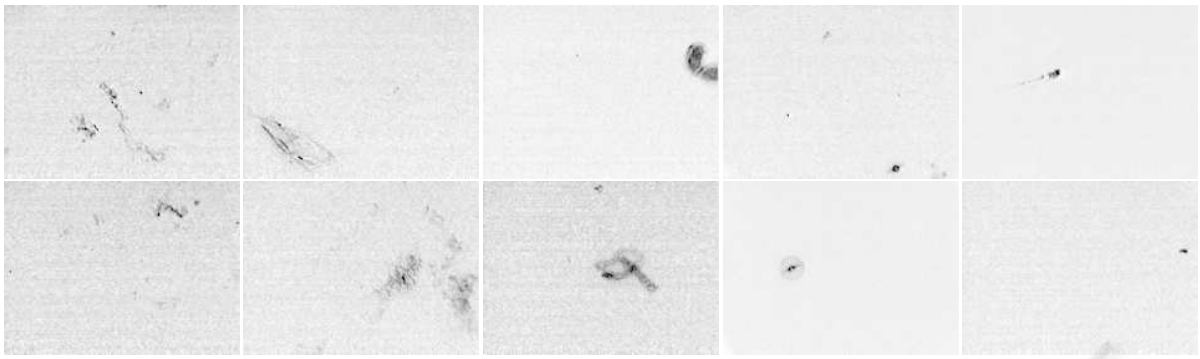

Figure 4: Sample images of our marine data: each column shows two images from each of the five classes, from left to right, algal aggregates, *Rhizosolenia* mats, chain colonial radiolarians, single sphere colonial radiolarians, and marine snow.

We measure information content using entropy. High entropy signals a heterogeneous field, an indication of activities. These patches likely contain at least a partial object. A patch with low entropy appears as a homogeneous field, containing no information on either objects, background, or their border relations. Therefore, it is safe to exclude low-entropy patches from a feature set. We eliminate 30% of patches with the lowest entropy and use the rest as actual templates.

4.3 Implementations

Our Matlab implementation is based on (Serre et al., 2005), in which preprocessing includes resizing every image so its height is 140 pixels and its aspect ratio is preserved. S1 and C1 parameters are left unchanged. Training patches are extracted from C1 maps at the second to the finest scale of Gabor filters. Patches are square; 4, 8, 12, and 16 pixels. From a training image, 32 patches are randomly selected, 8 of each size.

A data set is randomly divided into disjoint sets for training and testing. There are 30 training images from each class. The rest of the images from a class with less than 80 images form a test set. Otherwise, 50 test images are chosen. Accuracies are averages over 25 runs, all of which have different sets of randomly chosen training and testing images. The errors indicate 95% confidence interval of the mean values.

5 MARINE DATA SETS

The Pacific VPR imagery was collected in the North Pacific Subtropical Gyre during an August-September 2003 research cruise, which was part of NSF-funded projects to examine oceanic nitrogen cycling as influenced by marine diatom and algal aggregate distribution (Pilskaln et al., 2005). The slowly towed VPR was lowered 150m from the surface. This underwater video microscope system captured images at 60Hz. An interlaced scan pattern with odd/even fields is used. Due to rapid movements of both the imaging system and objects imaged, raw individual fields are interpolated. After post-processing, images are resized to a resolution of 640×480 pixels.

Table 1: Classification accuracies and running times. Each runtime is for a single experiment including on-line feature computations and classifications, but excluding off-line preprocessing steps; image segmentation for the traditional approach and Gabor filtering for the biological approach.

Experiment Specific	Accuracy %	Time min
Traditional Approach	76.69	40.83
The Standard Model		
no modification	78.16±1.11	251.44
patch overlap	79.96±1.16	149.13
overlap and entropy	80.73±1.03	116.88

The data set contains 488 images of marine organisms hand-selected and hand-labeled by oceanographer experts; examples are shown in Figure 4. Two of the 5 classes are phytoplankton: algal aggregates and *Rhizosolenia* mats. Two are protozoan zooplankton: chain colonial radiolarians and single sphere colonial radiolarians. The fifth class is marine snow (organic detritus). Phyto- and zooplankton are living plant and animal that drift in the water. Detritus are non-living particles sinking to the bottom layers of the ocean.

6 RESULTS AND DISCUSSIONS

Classification accuracies along with running times of the experiments are reported in Table 1. Comparing the traditional approach with the biological model and its variants, the former has much lower computational cost, allowing images to be categorized more quickly. Both approaches accomplish classification accuracies between 76-80%. Given a 75-80% labeling agree-

ment between expert biologists and 71-75% accuracies from similar works by (Blaschko et al., 2005) and (Lisin, 2006), our techniques perform very well on this very complex marine data set.

We compare the performance of the Standard Model with no modification versus when the number of extracted training patches is limited. The latter reduces running time by over 40%. Not only is the computational cost significantly lower, but also the classification results improve (using t-statistic confirms a significant difference at a 95% confidence level). These results suggest that, without modification, overlap among patches contains redundant information and decreases the separability among classes.

Measuring the information content of patches using entropy effectively removes patches that contain little to no information relevant to a categorization. The difference between the two accuracies, both with overlapping patches constraint applied, but with and without the use of entropy to select patches, show no statistical significance. The system performs equally well with 30% less features (without lowest-entropy training patches), a significant saving of a computational cost. The running time with the entropy selection process is reduced by over 20% compared to the system just using constrained patch overlap and by over 50% compared to the original Standard Model.

7 CONCLUSIONS

The potential of an artificial vision system based on biological principles is shown to be quite promising. A crucial advantage is its independence of image segmentation, a potentially high complexity processing step. When experimenting with the Standard Model, we were able to apply it, without re-tuning, and obtain accuracies as good as or better than a traditional approach, indicating good generalization capability.

To our knowledge, this is the first study that applied a biologically inspired model to a difficult underwater imaging domain. Our experimental results offer great promise for the analysis of large marine image data sets collected from unique open-ocean ecosystems. Automatic identification and quantification of the plankton and particle components, coupled with chemical and taxonomic composition analysis, will facilitate the production of a refined carbon and nitrogen budgets for this vast region, significant to our understanding of how climate change affects the dynamics of ocean ecosystems.

ACKNOWLEDGEMENTS

This work was supported by NSF (OCE-0325167).

REFERENCES

Benfield, M. et al. (2007). RAPID: Research on automated plankton identification oceanography. *Oceanography*.

Blaschko, M. et al. (2005). Automatic in situ identification of plankton. In *Proc. IEEE WACV*, pages 79–86.

Blu, T. and Unser, M. (2003). A complete family of scaling functions: The (α, τ)-fractional splines. In *Proc. IEEE Intl. Conf. Acoustics, Speech, and Signal Processing*.

Byun, J. et al. (2006). Automated tool for the detection of cell nuclei in digital microscopic images: Application to retinal images. *Molecular Vision*, 12:949–960.

Culverhouse, P. et al. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, 247.

De Valois, R. L. and De Valois, K. K. (1988). *Spatial Vision*.

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804.

Hays, G. C., Richardson, A. J., and Robinson, C. (2005). Climate change and marine plankton. *Trends in Ecology and Evolution*, 20(6):337–344.

Honjo, S., Manganini, S., Krishfield, R., and Francois, R. (2008). Particulate organic carbon fluxes to the ocean interior and factors controlling the biological pump: A synthesis of global sediment trap programs since 1983. *Progress In Oceanography*, 76(3).

Hu, M. (1962). Visual pattern recognition by moment invariants. *IEEE Trans. Information Theory*, 8:179–187.

Hubel, D. and Wiesel, T. (1959). Receptive fields of single neurons in the cat's striate cortex. *J. of Physiology*.

Lisin, D. A. (2006). *Image Classification with Bags of Local Features*. PhD thesis, Univ. of Mass., Amherst, MA.

Ojala, T., Pietikäinen, M., and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE PAMI*.

Pilskaln, C. et al. (2005). High concentrations of marine snow and diatom algal mats in the north pacific subtropical gyre: Implications for carbon and nitrogen cycles in the oligotrophic ocean. *Deep Sea Research*.

Ponce, J. et al. (2006). Dataset issues in object recognition. In *Towards Category-Level Object Recognition*.

Ravela, S. (2003). *On Multi-Scale Differential Features and their Representations for Image Retrieval and Recognition*. PhD thesis, Univ. of Mass., Amherst, MA.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.*

Serre, T., Wolf, L., and Poggio, T. (2005). Object recognition with features inspired by visual cortex. In *Proc. IEEE CVPR*. http://cbcl.mit.edu/software-datasets.

Sieracki, C., Sierackia, M., and Yentsch, C. (1998). An imaging-in-flow system for automated analysis of marine microplankton. *Marine Ecology Progress Series*.