Editorial

OPEN ACCESS

Biological Oceanography Needs New Tools to Automate Sample Analysis

P.F Culverhouse*

Centre for Robotics & Neural Systems, Plymouth University, Plymouth PL4 8AA

*Corresponding Author: Philip F Culverhouse, Centre for Robotics & Neural Systems, Plymouth University, Plymouth PL4 8AA. Tel: +44 (0) 1752 600600; E-mail: pculverhouse@plymouth.ac.uk

Editorial

What do an oil spill, invasive species and routine fish stock surveys have in common? They all have to use an expert to manually identify and count the species involved. Changes occur in plankton distributions following an oil spill, invasive species introduced through discharge of ship's ballast water can seriously impact on ecosystems and fish larval identification must be carried out for routine stock assessments of a fishery. These are all examples where the necessary expertise requires years of training before competence is achieved.

There has been a serious decline in scientists who have this biological expertise over the last 50 years. Training of taxonomists and specialized training programs has been cut back in Universities and Research Laboratories[1]. As a result there are only about 4,000–6,000 professional taxonomists worldwide and the demand for skilled analysts is escalating. Therefore, there exists a bottleneck[2] in the identification of natural organisms due to the lack of qualified taxonomists and analysts across a very wide range of taxa from agriculture, marine sciences, aquaculture and border control inspection to paleontological analysis of oil-bearing rocks.

Computer vision and artificial intelligence offers a way of reducing the burden on these skilled analysts. Advances over the past decade in computer-based identification mean that some classes of identification are possible without human operator intervention in routine analysis. The steps involved are broadly: first obtain a set of specimen images that can be used for training and testing purposes (an expert must identify and label each specimen with a high degree of accuracy). Then use computer vision techniques to extract meaningful features from the set of images. Then by using machine learning techniques, patterns in the data can stored and used for later identification of previously unseen images. Performance on training data and test sets can approach 90%, when individual specimens are present in an image form. Performance in cluttered scenes is often poor, but progress is being made.

Typically objects are detected in each image through a monochrome conversion and features processed for numerical descriptors of invariant moments, contour shapes and grey level histograms. A numerical vector containing all the descriptors is then assembled for each specimen image. Sets of these are fed into machine learning methods such as support vector machine[3] and random forest[4]. All current methods require at least 40 examples of each class. Repeated randomized training continues until a minimum learning performance is achieved on both the test and the training set. The trained classifier is then evaluated on previously unseen data and applied to field collected specimens for example[5-17]. A serious shortcoming of current automatic methods is the poor performance at identifying rare organisms.

It is important that the biological oceanography community embrace these new techniques for specimen identification, in the same way that the geophysical oceanographers have welcomed instrumentation, and benefitted from widespread acceptance of Conductivity Temperature and Depth (CTD) measurements that have revolutionized our understanding of the physical dynamics of our seas and oceans. However, there is much resistance to new methods in the biological and ecological community, as these new tools are not refined enough to replace plankton analysts per se, but they can assist people to increase analytical throughput.

What is often not understood is that humans suffer from cognitive shortcomings that limit their performance in identification tasks. A number of studies led by Culverhouse and others[18-24,26-28] reveal that plankton identification by experts varies from better than 90% to worse than chance, according to the cognitive biases extent at the time of the activity. Boredom and exhaustion are significant factors affecting performance, so too positivity bias and recency effects[25]. Day to day repeatability is also in question, with 14 out of 21 experts in one study[26] performed as experts, being self consistent to better than 90%, the remainder ranged from 80% down to less than 20% repeatability. Experts often work alone, with no quality assurance or cross checking. At the present time, machine performance is subject to closer scrutiny than expert performance.

References

1. The global Taxonomy initiative. (2015) Convention on Biological Diversity.

2. MacLeod, N., Benfield, M., Culverhouse, P. Time to Automate Identification. (2010) Nature 467(7213): 154-155.

3. Vapnik, V.N., Kotz, Samuel. Estimation of Dependences Based on Empirical Data. (2006) Springer.

4. Breiman, L. Random Forests. (2001) Machine Learning 45(1): 5-32.

5. Remsen, A., Samson, S., Hopkins, T. What you see is not what you catch: a comparison of concurrently collected net, Optical Plankton Counter, and Shadowed Image Particle Profiling Evaluation Recorder data from the northeast Gulf of Mexico. (2004) Deep-Sea Research I 51: 129-151.

6. Cowen, R.K., Guigand, C.M. In situ Ichthyoplankton Imaging System (ISIIS): system design and preliminary results. (2008) Limnol Oceanogr Methods 6(2): 126-132.

7. Sieracki, C.K., Sieracki, M.E., Yentsch, C.S. An imaging-in-flow system for automated analysis of marine microplankton. (1998) Mar Ecol Prog Ser 168: 285-296.

8. Picheral, M., Guidi, L., Stemmann, L., et al. The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. (2010) Limnol Oceanogr Methods 8(9): 462–473.

9. Davis, C.S., Gallager, S.M., Berman, M.S., et al. The Video Plankton Recorder (VPR): Design and initial results. (1992) Arch Hydrobiol Beih 36: 67-81.

10. Culverhouse, P.F., Williams, R., Simpson, R., et al. HAB-BUOY: A new instrument for monitoring HAB species. XI International Conference on Harmful Algae, Cape Town, South Africa, 15-19 November 2004. (2006) Afr J Mar Sci 28(2): 245–250

11. Gorsky, G., Ohman, M.D., Picheral, M., et al. Digital zooplankton image analysis using the ZooScan integrated system. (2010) J Plank Res 32(3): 285-303

12. Bell, J.L., Hopcroft, R.R. Assessment of ZooImage as a tool for the classification of zooplankton. (2008) J Plank Res 30(12): 1351-1367.

13. Hu, Q., Davis, C. Accurate Automatic Quantification of Taxa-Specific Plankton Abundance Using Dual Classification with Correction. (2006) Mar Ecol Progr Ser 306: 51-61.

14. Shortis, M.R., Ravanbakskh, M., Shafait, F., et al. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. (2013) Videometrics, Range Imaging, and Applications XII, and Automated Visual Inspection, SPIE 8791: 0G.

15. Zion, B., Shklyar, A., Karplus, I. Sorting fish by computer vision. (1999) Computers and Electronics in Agriculture 23(3): 175-187.

16. Rova, A., Mori, G., Dill, L.M. One Fish, Two Fish, Butterfish, Trumpeter: Recognizing fish in underwater video. (2007) IAPR Conference on Machine Vision Applications 404-407.

17. Bi, H., Cook, S., Yu, H., et al. Deployment of an imaging system to investigate fine-scale spatial distribution of early-life stages of the ctenophore Mnemiopsis leidyi in Chesapeake Bay. (2012) J Plank Res 1-11.

18. Simpson, R., Williams, R., Ellis, R., et al. Biological pattern recognition by neural networks (1992) Mar Ecol Prog Ser 79: 303–308.

19. Culverhouse, P.F., Ellis, R., Simpson, R.G., et al. Automatic categorisation of 5 species of Cymatocylis (Protozoa, Tintinnidae) by artificial neural network. (1994) Marine Ecology Progress Series 107: 273–280.

20. Culverhouse, P.F., Williams, R., Reguera, B., et al., Automatic classification of field-collected dinoflagellates by artificial neural network. (1996) Mar Ecol Prog Ser 139: 281–287.

21. Ellis, R., Simpson, R., Culverhouse, P.F., et al. Committees, collectives and individuals: Expert visual classification by neural network. (1997) Neural Computing and Applications 5(2): 99–105.

22. Ginsburg, R.N.. Perspectives on the blind test. (1997) Marine Micropaleontology 29(2):101-103.

23. Kelly, M.G. Use of similarity measures for quality control of benthic diatom samples. (2001) Water Res 35(11): 2784-2788.

24. Culverhouse, P.F., Williams, R., Reguera, B., et al. Do Experts Make Mistakes? A comparison of human and machine indentification of dinoflagellates. (2003) Mar Ecol Prog Ser 247: 17-25.

25. Culverhouse, P.F. Human and machine factors in algae monitoring performance. (2007) Ecological Informatics 2(4): 361-366.

26. First, M.R., Drake, L.A. Performance of the human "counting machine": evaluation of manual microscopy for enumerating plankton. (2012) Journal of Plankton Research 34(12): 1028-1041.

27. Evans, J.St.BT. Bias in Human Reasoning: Causes and Consequences. (1989) Hove Laurence Erlbaum Associates 125.

28. Culverhouse, P.F., MacLeod, N., Williams, R., et al. An empirical assessment of the consistency of taxonomic identifications. (2010) Marine Biology Research 10(1): 73-84.