Classification of Seafloor Habitats using Genetic Programming

Sara Silva
Centre for Informatics and Systems
University of Coimbra
Polo II, 3030 Coimbra, Portugal
sara@dei.uc.pt

Yao-Ting Tseng
Centre for Marine Science and Technology
Curtin University of Technology
GPO Box U1987, Perth, WA 6845, Australia
y.tseng@cmst.curtin.edu.au

ABSTRACT

In this paper we use Genetic Programming for the classification of different seafloor habitats, based on the acoustic backscatter data from an echo sounder. By developing a different fitness function and dividing the multiple-class problem into several two-class problems, we were able to improve the results presented in a previously published work, providing a better discrimination between most of the seafloor types used in this study. We discuss the quality of these results and provide ideas to further improve the classification performance.

1. INTRODUCTION

Genetic Programming (GP) can solve complex problems by evolving computer programs using Darwinian evolution and Mendelian genetics as sources of inspiration [1,2]. Many GP systems represent the programs as trees. Tree-based GP is the most widely used, but its nature does not make it particularly suited for multiclass classification tasks, although some studies have already been developed on this subject [3–5].

The aim of this work is to provide a better understanding of the acoustic backscatter from marine macro-benthos (MMB), including mainly seagrass, algae, and other marine organisms living on the seafloor. Since these organisms live on or around their substrates, the understanding of the acoustic backscatter from their substrates is also essential.

The analysis of the acoustic backscattered signals of MMB and related substrates has been studied with a variety of different approaches [6–9]. One of them [9] has been the target for improvement in a work using GP [10], where GP was able to provide an improved discrimination between the different seafloor habitats. The initial motivation to use GP for this task came from a work on diesel engine diagnosis [11]. Other works on fault diagnosis using GP are available [12, 13].

GECCO'05 Late-Breaking Papers, June 25–29, 2005, Washington, DC, USA.

Here we tackle the same problem studied in [10], developing a new fitness function for the GP system, testing the pairwise separability of the different classes involved in the study, and ultimately dividing the 5-class problem into several 2-class problems, whose solutions can be joined to provide a perfect discrimination of most of the seafloor habitats.

The next section describes the data used in this study, how it was acquired and prepared for being used. Section 3 describes the GP system used, its main parameters and the fitness function developed for this particular problem. Section 4 describes the results achieved, and how they were combined to build the final solution. Section 5 discusses the quality and usefulness of the proposed solution, suggesting future developments of this work. Finally, Section 6 concludes this study.

2. THE DATA

This section describes the data collection process, the removal of incomplete data and definition of representative data sets, and the statistical preprocessing suffered by the data before being used by the GP system.

2.1 Data Acquisition

The acoustic backscattered signals were collected from Cockburn Sound Western Australia on the 10th of August 2004 by a SIMRAD EQ60 single beam echo sounder. The data collection was made on two sites of shallow coastal waters where the water depths were less than 6 meters. In site 1, the seafloor habitats are predominantly sand, seagrass 1 (*Posidonia sinuosa*), and seagrass 2 (*Posidonia australis*). On the other hand, site 2 mainly consists of sand, reef and macro algae with canopy heights much higher than both of the seagrasses in site 1. Along with the collection of the acoustic data, synchronized tridimensional (3D) still images were also taken simultaneously.

Figure 1 shows an echo sounder transmitting a signal to the seafloor. The sound is backscattered from the seafloor to the sea surface, and back to the seafloor, several times. The echo sounder receives several returns for each sample. Figure 2 represents a typical sample of acoustic backscatter collected from an echo sounder, showing several echo returns. Although SIMRAD EQ60 provides both 38 and 200kHz sampling ability, only the 200kHz signals have been used in our study, due to its higher resolution of $25\mu s$ sampling rate. The volume backscatter coefficient (in decibel scale) was used due to its ready availability.

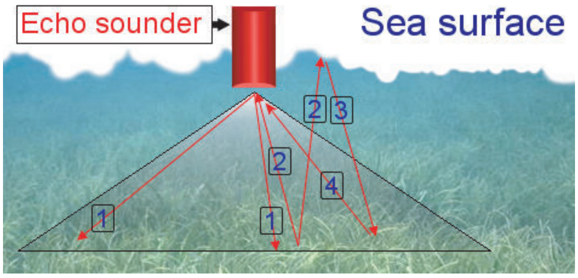

Figure 1: Sound transmitted from the echo sounder to the seafloor (1), echo from the seafloor producing the first bottom return (2), echo from the sea surface to the seafloor (3), and again from the seafloor to the sea surface and to the echo sounder, producing the second bottom return (4).

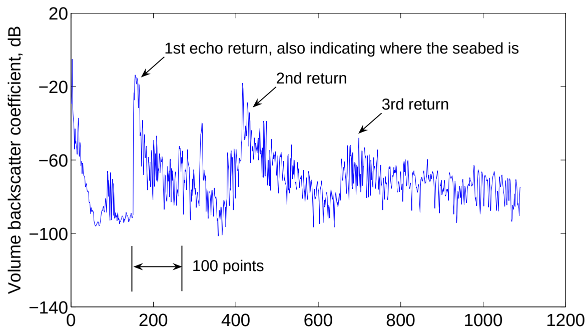

Figure 2: A typical sample of the acoustic backscatter collected from an echo sounder. Several echo returns are present.

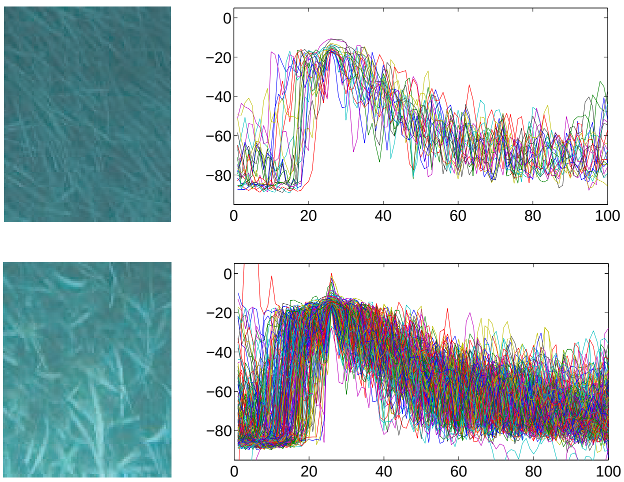

Figure 3: Still images (left) and echo signals (right) from both seagrasses: *Posidonia sinuosa*(top) and *Posidonia australis*(bottom).

2.2 Defining Data Sets

Due to the cost and technical limitations of this study, only 1232 samples of both echoes and still images were acquired. The MMB and the related substrates were roughly classified into five classes: sand, bare reef, macro algae, seagrass 1 (*P. sinuosa*), and seagrass 2 (*P. australis*) according to the visual interpretation of the 3D still images. From now on, we will refer to the seagrasses simply as sinuosa and australis. After further examination of the images and the echoes, 689 samples were rejected for not being fully intact, and some others were discarded for containing mixed habitat types. In the end, 300 samples were used as pure representatives of the five classes, unevenly distributed like this: 81 (sand), 10 (reef), 8 (algae), 21 (sinuosa), 180 (australis). Each sample consisting of several bottom returns was then truncated to contain only the first bottom return (see Figure 2), represented by a 100-point sequence that is believed to fully describe the interactions between the transmitted sound and the respective targets.

Figure 3 shows an example of still images of both seagrasses and the available echoes (after being truncated) for these classes. The intra-class variety and inter-class similarity of the echo signals immediately hints at the difficulty of this problem.

2.3 Statistical Preprocessing

Before being given to the GP system, each sample suffers a major transformation, one that may well determine the success or failure of the GP learning. Each of the 100-point sequences ($S = \{p_1, p_2, ..., p_n\}, n = 100$) is reduced to only seven statistical features ($F = \{x_1, x_2, ..., x_7\}$):

Kurtosis: $x_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} p_i^4}{x_4^2}$

Maximum: $x_2 = \max p_i, i = 1, 2, ..., n$

Mean: $x_3 = \frac{1}{n}\sum_{i=1}^{n} p_i$

Second-order Moment from Origin: $x_4 = \frac{1}{n}\sum_{i=1}^{n} p_i^2$

Skewness: $x_5 = \frac{\frac{1}{n}\sum_{i=1}^{n} p_i^3}{x_4^{3/2}}$

Standard Deviation: $x_6 = (\frac{1}{n-1}\sum_{i=1}^{n}(p_i - x_3)^2)^{\frac{1}{2}}$

Minimum: $x_7 = \min p_i, i = 1, 2, ..., n$

This particular set of statistical features was based on the set by [11]. After calculating the different statistical features for all the samples, the 300-element vectors obtained for each feature are normalized. Each fitness case of the GP system is a 7-tuple with all the statistical features of the sample ($x_1, x_2, ..., x_7$), along with an identifier of the class to which it belongs.

By themselves, these seven features do not seem to have the ability to discriminate between our several classes. Figure 4 shows the plotting of four of these features (Maximum, Mean, Skewness, Standard deviation) for the five classes involved. The plots produced by the remaining features are not very different from the ones shown. The samples of the least represented classes were randomly replicated until each class contained 180 points in the plot, for visualization purposes only. For an easier recognition of the ranges occupied by each classes, consecutive points are connected by

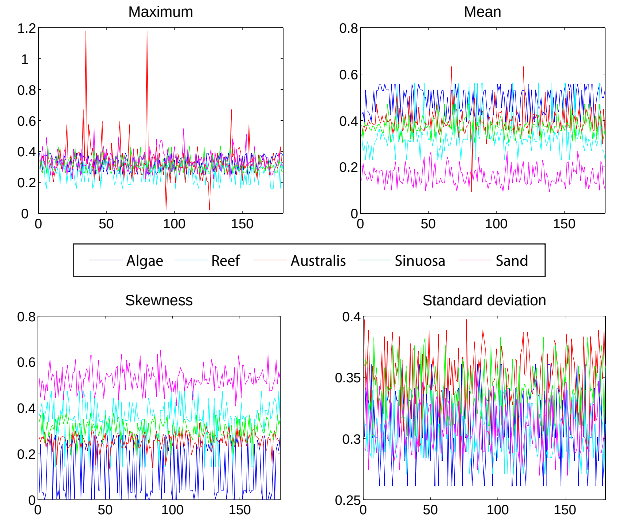

Figure 4: Plotting of four statistical features for the five classes involved in the study.

Table 1: Main running parameters of the GP system used.

Function Set	$\{+, -, \times, \div\}$ (protected [1])
Terminal Set	$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$
Population Initialization	Ramped Half-and-Half [1]
Population Size	500
Maximum Tree Depth	initial: 6, final: unlimited [15]
Operator Rates	crossover/mutation: 0.5/0.5
Reproduction Rate	0.1
Selection for reproduction	tournament [16], size 50
Selection for survival	replacement (no elitism)

lines. Sand is the class that seems to be more easily separated from the rest, but still none of the single features is able to do that. GP is expected to be able to combine the single features into a compound feature that will avoid the overlapping between the ranges of any two classes.

3. THE GP SYSTEM

This section describes the main parameters of the GP system used, as well as the fitness function developed for this work.

3.1 Parameters

As in [10], the GP system used was an adaptation of GPLAB, a GP toolbox for MATLAB [14], with the main running parameters indicated in Table 1.

The function set used was shorter than in [10], containing only the most basic operators. As in [10], we have used the Lexicographic Parsimony Pressure tournament [16] and the Heavy Dynamic Limit [15] on tree depth to avoid excessive code growth, but without using any traditional static limit.

3.2 Fitness function

The most influential element of any GP system is the fitness function. After studying the limitations of the system

```
nclasses = number of classes
nsamples_c = number of samples in class c
min_c = minimum value plotted in class c
max_c = maximum value plotted in class c

overlapped = 0
for c = 1 to nclasses
   for s = 1 to nsamples_c
      value_s = value plotted for sample s
      for nc = 1 to nclasses, nc <> c
         if value_s between min_nc and max_nc
            overlapped = overlapped + 1/nsamples_c

fitness = 100 * overlapped / nclasses
```

Figure 5: Pseudocode of the fitness function.

presented in [10], we have developed a new fitness function that allows a better learning of the feature combinations necessary to differentiate the different classes involved.

The previous fitness function [10] was based on inter-class and intra-class distances, inspired by the fitness function on [11]. But it lacked the full objectivity needed to guide GP through a difficult error landscape. This allowed the GP system to "cheat", by dispersing the values plotted (like in Figure 4) by the compound feature (the candidate solution) such that the calculated fitness value was indeed increased, but without really improving the separability between the classes.

The new fitness function is not presented as a neat formula, but as a pseudocode procedure that totally disregards the notion of distances between and within classes. Instead, it simply calculates the percentage of points in the plot (like in Figure 4) that fall within the range of more than one class. Minimizing this percentage is our final goal, a simple and clear objective that allows the GP system complete freedom to devise any possible discrimination strategy, as long as it reaches its purpose. Figure 5 shows the pseudocode for the fitness function. Because the available samples are unequally distributed between the classes, we give more weight to the points of the under-represented classes, such that each class contributes equally to the calculation of the fitness value. So, a point from class algae (8 samples) weights 22.5 times more than a point from class australis (180 samples).

4. RESULTS

This section shows two types of results. First we present some preliminary tests regarding the pairwise separability of the classes involved. We also show additional tests where we have tried to separate more than two classes at the same time. Then we propose a solution for discriminating between most of the classes involved in this study, by dividing the 5-class problem into several 2-class problems.

4.1 Pairwise Separability

Our first approach to solve the multiclass classification problem was to check the separability between all pairs of classes.

4.1.1 Pairing with Sand

Just by looking at Figure 4, one can immediately see that Sand can be separated from Algae, Sinuosa and Australis using the single features x_1 (Kurtosis, not shown in the figure), x_3 (Mean) or x_5 (Skewness). Reef was the only class that could not be completely separated from Sand using single features. So we performed a few GP runs, using the system described in Section 3, and easily found a couple of compound features, short and simple, that can do the job: $x_2 - x_3$ and $x_3 - x_2$. We hypothesize that a good way to attack the 5-class problem would be to first separate Sand from the remaining group of classes, and then proceed to solve the remaining 4-class problem.

4.1.2 Pairing with Reef

We proceeded with the pairwise separability tests by trying to separate Reef from the remaining classes (except Sand, of course). Finding a compound feature that perfectly discriminates between Reef and Algae proved to be a fairly simple task for the GP system:

$$\frac{x_1 x_2}{x_3(x_3 - x_6)}$$

Discriminating between Reef and Sinuosa also posed no difficulties:

$$\frac{x_4}{x_6(x_7 - 1)} - x_1$$

On the other hand, separating Reef from Australis proved to be a difficult task. In most of the runs, the GP system converged to compound features that could not separate both classes perfectly, meaning that the fitness (weighted percentage of overlapped points, see Section 3.2) did not reach zero. But eventually it was able to find a perfect solution:

$$\frac{x_3(x_3 x_5 - x_6)(x_5 + x_6 - x_7)}{x_5(x_1 - 2x_2 + x_5 - x_6)} - x_5$$

Due to the difficulties mentioned, we hypothesize that separating Reef from all the remaining classes may not be an easy task.

4.1.3 Pairing with Algae

To continue testing the pairwise separability, we tried to separate Algae from the remaining classes, both seagrasses. Finding a compound feature to discriminate between Algae and Sinuosa was very easy: $x_4 + x_5 + x_6$. However, separating Algae from Australis proved to be very difficult. The GP system did not converge to any perfect solution in a reasonable amount of time, only finding almost perfect compound features that could not completely discriminate among both classes. We hypothesize that it may be very difficult or impossible to separate Algae from a set of classes that includes Australis.

4.1.4 Pairing the Seagrasses

To finish the tests on pairwise separability, we have finally tried to separate both seagrasses, but the GP system could not find any perfect solution. In fact, the fitness of the solutions it converged to was far from zero. We present one of the best solutions found, with fitness 44.0 (44% of the samples are overlapped), in Section 4.3.4, and hypothesize that separating both seagrasses may be an impossible task using the current data, features and settings.

4.2 Separability of Multiple Classes

From the results achieved in Section 4.1 we proceeded to more complex tests where we tried to discriminate between more than two classes at the same time. The pairwise separability tests have revealed a group of classes with good prospects of being easily separated from each other: Sand, Algae, and Sinuosa. The separation between any two of these classes was achieved with short and easy to find compound features, so we decided to check if the GP system could find a compound feature capable of discriminating between the three. A short solution was found with not much difficulty (the plot is shown in Figure 6, left):

$$x_6 + 2\left(\frac{x_5}{x_3 + x_6}\right)$$

We tried the same thing with a slightly different group: Sand, Algae, and Reef. From the results of pairwise separability with Reef, we were expecting this to be a more complex task. Although with more difficulty, the GP system was once again able to find a perfect solution (the plot is shown in Figure 6, right):

$$\begin{aligned} x_7 &- \left(2x_1 + x_2 x_5 + \frac{x_5}{x_3}\right) \\ &- \frac{x_3 x_4 x_5}{x_2 x_3(x_1 + x_5 - x_7 + x_2 x_5) + x_5(x_2 + x_3 x_6)} \end{aligned}$$

In the plots of Figure 6 the samples of the least represented classes were randomly replicated until each class contained 81 points, the number of samples of the largest class in this case.

We did not perform any additional tests with other groups of three classes. Instead, we tried to separate a group of four classes, made of the classes contained in the two previous trios: Sand, Algae, Sinuosa, Reef. The GP system did not find any perfect solutions, so we moved to a different strategy.

4.3 Divide and Conquer

Because of the failure reported in Section 4.2 in discriminating among the four easiest classes of our problem, we adopted a divide and conquer strategy, dividing the 5-class problem into several 2-class problems. We used the pairwise separability results obtained in Section 4.1 to guide this process.

4.3.1 Separating Sand

As already stated, Sand was the easiest class to separate from any of the others, so we began by searching a compound feature to discriminate between Sand and the set of all the remaining classes. As expected, this was an easy task for our GP system:

$$\frac{x_3^2}{x_2} \tag{1}$$

The plot produced by this perfect solution is shown in Figure 7, top left. The dashed line indicates the boundary between Sand and the remaining classes.

4.3.2 Separating Reef

We proceeded by searching for a compound feature to discriminate between Reef and the set of remaining classes (except Sand, of course). Due to hard pairwise separability between Reef and Australis, we were expecting this to be a very difficult task, and so it was. Most of the GP runs converged to non perfect solutions, or to extremely complex

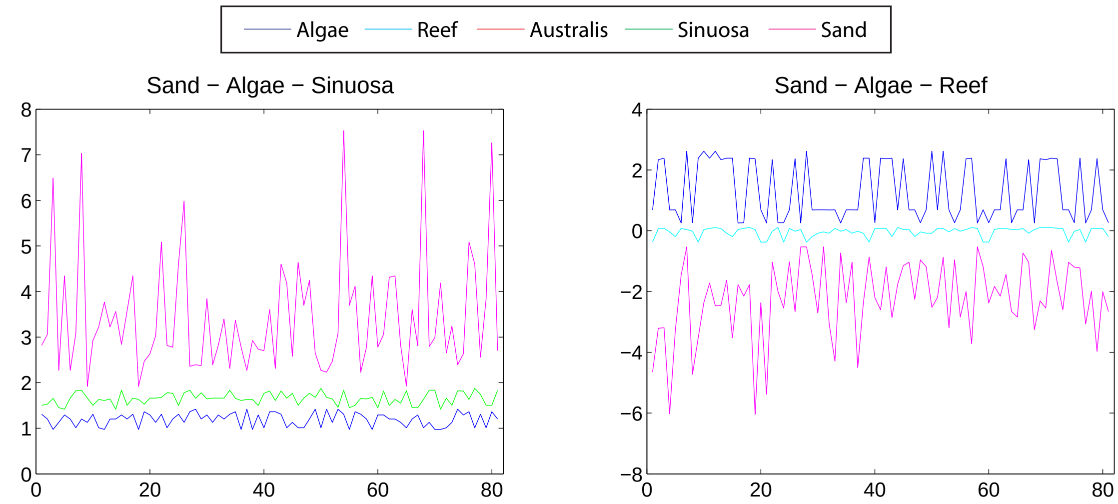

Figure 6: Plots produced by the compound features found for two 3-class problems.

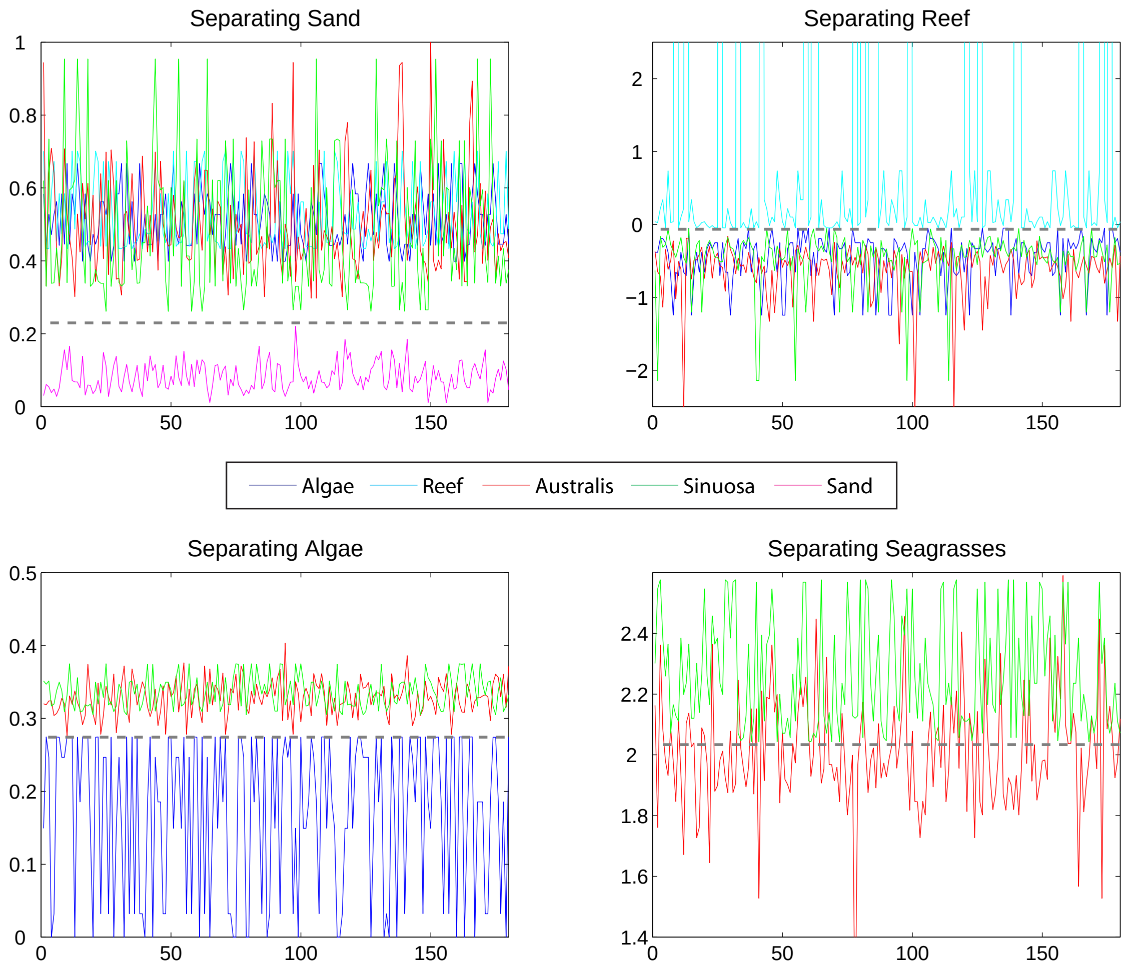

Figure 7: Plots produced by the compound features found for the several sub-problems: Separating Sand (top left, Section 4.3.1), Separating Reef (top right, Section 4.3.2), Separating Algae (bottom left, Section 4.3.3), Separating Seagrasses (bottom right, Section 4.3.4). The dashed lines represent the thresholds specified in Section 4.4.

compound features, as this one:

$$\frac{x_6(x_5 - x_7 - x_2x_6)}{x_3 - x_4x_6} \times \\ \times \left[1 + x_4x_5(x_3 - x_2) + x_3x_7(x_3 - x_4) - \frac{x_3^3x_7}{x_4x_5} - \right. \\ - \frac{x_2(x_3x_7 - x_5) + x_3(x_3x_7 + x_5)}{x_4} + \frac{x_3^2}{x_5} + \\ \left. + \frac{x_3x_5(x_3 - x_2) + x_4x_5^2(x_2 - x_3)}{x_4(x_1x_3x_4x_5 - x_4x_5(x_3 + 1) + x_3)}\right]^{-1} \tag{2}$$

The plot produced by this perfect solution is shown in Figure 7, top right. The dashed line indicates the boundary between Reef and the remaining classes.

4.3.3 Separating Algae

We continued our divide and conquer strategy by searching for a compound feature to discriminate between Algae and the two remaining classes, the seagrasses. We were not expecting to find a perfect solution for this problem, because we had never achieved the pairwise separation between Algae and Australis. We were wrong, and given enough persistence we were eventually able to find a perfect solution (that is obviously also a solution for the pairwise problem):

$$\frac{x_4x_5(x_1 + x_4)}{x_4 + x_5} \times \frac{x_6 + (x_3 + x_7)(2x_2 + x_5)}{2x_2 + x_5} \times \\ \times \left[\frac{x_6(x_2 + 2x_7)(x_6 + x_2x_5 + x_5^2(x_2 + x_5 + 1))}{x_6(x_2 + 2x_7) + x_5^2(x_2 + x_5)(x_2 + 2x_7)} \right. \\ \left. + \frac{x_5x_7(x_2 + x_5)(x_1 + x_3 - x_7)}{x_6(x_2 + 2x_7) + x_5^2(x_2 + x_5)(x_2 + 2x_7)}\right]^{-1} \tag{3}$$

The plot produced by this perfect solution is shown in Figure 7, bottom left. The dashed line indicates the boundary between Algae and the remaining classes.

4.3.4 Separating Seagrasses

After successfully separating the previous classes, we were left with the problem of discriminating between both seagrasses. This is the exact same problem already dealt with in Section 4.1.4, that we have concluded to be too difficult, if not impossible. Here we present one of the best solutions that could be found:

$$x_2 + x_4 + x_5 - \frac{2(x_6(x_3 - 1) + 2x_2 - x_1)(x_2 - x_1)}{x_4x_5x_6} + \\ + \frac{x_5x_7}{x_3(x_1 + x_3 + x_4)(x_6 - 2x_5)} - \frac{x_2(x_3 + x_6)}{x_1x_5} + \\ + \frac{x_1x_4x_6(x_5(x_6 - x_2) - x_3 + x_4)(x_6 - 2x_5)(x_7 + x_1x_3)}{(x_1 + x_2 - x_5)(x_2x_7 + x_1^2(x_6 - 2x_5)(x_7 + x_1x_3))} \tag{4}$$

The plot produced by this far from perfect solution is shown in Figure 7, bottom right. The dashed line indicates a possible boundary to approximate a rough discrimination between both seagrasses.

4.4 The proposed solution

The results of the divide and conquer strategy presented in Section 4.3 can now be joined together to produce a candidate solution for the original 5-class problem. This solution is a binary tree where each node compares a compound feature with a threshold value, in order to determine if a class is already identified or if other nodes need to be visited. The threshold values are the dashed lines plotted in Figure 7. Figure 8 shows the proposed solution.

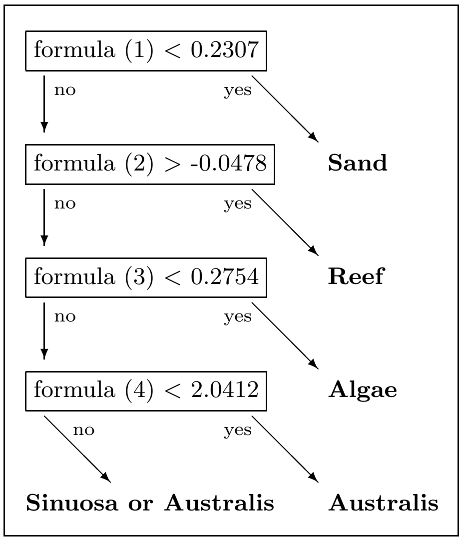

Figure 8: Proposed solution for the multiple class classification problem. Formulas (1), (2), (3) and (4) can be found in Section 4.3.

5. DISCUSSION AND FUTURE WORK

The results presented in Section 4 represent a major improvement over the results published earlier [10]. In the previous work, all the five classes were partially overlapped, while in this work only the seagrasses could not be completely separated from each other. The reasons for this limitation may be inherent to the data, that may not contain enough information to perform this discrimination, or to the large removal of information caused by the adoption of the statistical features as sole representatives of the samples. Using different statistical features, or ones that concentrate on smaller areas of the 100-point sequences, may allow us to finally distinguish between these two similar classes.

From the marine science point of view, it may not even be that important to achieve a solution that discriminates among all the classes. For example, it is known that live seagrasses only exist on sand, and algae only grow on reef. So the chances of erroneously identifying seagrass as algae are very low when most of the seafloor is sand. Likewise, one will hardly misidentify algae as seagrass when the substrates are reef. With this knowledge, it may be possible to reduce the complexity of the problem and concentrate the efforts on solving only the most important and practical issues.

In spite of its apparent quality, the solution proposed must be used with caution. Because the available data samples were highly unbalanced between the different classes, it was not possible to perform any cross-validation of the results. The solution may be biased toward some outliers that may be present and may represent a high proportion of data in the most under-represented classes.

New data is being collected that will allow us to test the proposed solution more thoroughly. If the new results turn out to be poor, then we are probably facing the problem of overfitting, a phenomenon that is already hinted by the large size of some of the compound features found. If the new data confirms this, it will also serve to build a new and more robust solution that can avoid this problem.

It should also be noted that we have only used pure habitats types to derive the proposed solution. If we had used mixed types, finding compound features to completely separate between classes would not only be virtually impossible, but would also not make much sense in practical terms. The fact is that the proposed solution is not appropriate for dealing with mixed types at all. Since the divide and conquer strategy seems to be a promising way of dealing with the inadequacy of GP for multiclass classification problems, in the future we may adopt a different technique for solving each of the sub-problems.

Like what typically happens when training an artificial neural network, GP can also be taught to output a number between 0 and 1 that can be interpreted as representing the likelihood that a given sample belongs to a given class. Joining the solutions of the different sub-problems would then result in a vector containing as many elements (numbers between 0 and 1) as classes involved in the study. This would allow the GP system to perform a fuzzy classification, as opposed to the sharp and clear-cut classification performed by the current solution, which hardly represents the real conditions of most samples collected in the natural environment.

Finally, the fitness function used in this work, although better than the one previously published [10], is still lacking an important feature. It can lead the GP system to the complete separation of classes, but once it gets there it does not promote any further separation, meaning it does not reward the solutions that present a larger distance between the classes. The result is that there is only a thin range of values from where to choose the thresholds of the final solution, something that will probably impair the performance in new data sets. The current fitness function can be extended to promote a larger distance between classes. Currently, it assumes values between 100 and 0, where the null value corresponds to the best cases, with no superposition of classes. In the future, the range of possible fitness values may reach below 0, where lower values represent cases with no superpositions and larger distances between classes, the truly ideal situation. But there are, of course, many other possibilities for improving the fitness function and the general performance of the entire GP system.

6. CONCLUSIONS

In this paper we have illustrated the usage of Genetic Programming on the classification of seafloor habitats. Using a fitness function different from the previous one [10] and dividing the multiple-class problem into several easier two-class problems, we have proposed a solution that represents a major improvement over the results published earlier. From the five classes involved in our classification problem, only two could not be completely separated from each other. The quality of these results is a motivation for performing further validation with new data, and for developing other GP techniques appropriate for solving harder problems of seafloor habitat classification.

7. ACKNOWLEDGEMENTS

This work was partially supported by Fundação para a Ciência e a Tecnologia, Portugal (SFRH/BD/14167/2003). The authors acknowledge their PhD supervisors: Ernesto Costa from University of Coimbra, Portugal; Alexander Gavrilov and Alec Duncan from Curtin University of Technology, Australia. Thank you also to Michael Harwerth from Germany, for many fruitful discussions and suggestions regarding this work. Great thanks go to the 3-D still image system leader, Andrew Woods. The data collection was funded by Cooperative Research Centre for Coastal Zone, Estuary & Waterway Management. The construction of the synchronized 3-D still image system with the SIMRAD EQ60 echo sounder was mainly done in the Centre for Marine Science and Technology based at Curtin University of Technology under the Epi-benthic Scattering Project (ESP), which is a subproject of a bigger Coastal Water Habitat Mapping project. Thanks go to all ESP members, especially project leader Rob McCauley and project manager John Penrose.

8. REFERENCES

[1] J. R. Koza. *Genetic programming - on the programming of computers by means of natural selection*. The MIT Press, Cambridge, Massachusetts, USA, 1992.

[2] W. Banzhaf, P. Nordin, R. E. Keller, and F. D. Francone. *Genetic programming - an introduction*. Morgan Kaufmann, San Francisco, 1998.

[3] W. Smart and M. Zhang. Probability based genetic programming for multiclass object classification. In C. Zhang, H. W. Guesgen, and W.-K. Yeap, editors, *PRICAI-2004*, LNCS, 251–261, Auckland, New Zealand, 2004. Springer.

[4] Y. Zhang and M. Zhang. A multiple-output program tree structure in genetic programming. In R. McKay, editor, *Proceedings of 2004 Asia-Pacific Workshop on Genetic Programming*, Cairns, Australia, 2004. Springer.

[5] W. Smart and M. Zhang. Using genetic programming for multiclass classification by simultaneously solving component binary classification problems. In M. Keijzer, A. Tettamanzi, P. Collet, *et al.*, editors, *EuroGP-2005*, LNCS, 227–239, Lausanne, Switzerland, 2005. Springer.

[6] J. Tegowski, N. Gorska, and Z. Klusek. Statistical analysis of acoustic echoes from underwater meadows in the eutrophic puck bay (southern baltic sea). *Aquatic Living Resources*, 16:215–221, 2003.

[7] B. Chakraborty, V. Mahale, C. de Sousa, P. Das. Seafloor classification using echo-waveforms: a method employing hybrid neural network architecture. *IEEE Geoscience and Remote Sensing Letters*, 1(3):196–200, 2004.

[8] D. Li, M. R. Azimi-Sadjadi, and M. Robinson. Comparison of different classification algorithms for underwater target discrimination. *IEEE Transactions on Neural Networks*, 15(1):189–194, 2004.

[9] P. J. W. Siwabessy, Y.-T. Tseng, A. N. Gavrilov. Seabed habitat mapping in coastal waters using a normal incident acoustic technique. In *ACOUSTICS-2004*, Gold Coast, Australia, 2004. Australian Acoustical Society.

[10] Y.-T. Tseng, A. N. Gavrilov, A. J. Duncan, M. Harwerth, S. Silva. Implementation of genetic programming toward the improvement of acoustic classification performance for different seafloor habitats. In *Oceans'05 Europe*, Brest, France, 2005. IEEE Press.

[11] R. Sun, F. Tsung, and L. Qu. Combining bootstrap and genetic programming for feature discovery in diesel engine diagnosis. *International Journal of Industrial Engineering*, 11(3):273–281, 2004.

[12] P. Chen, M. Taniguchi, T. Toyota, and Z. He. Fault diagnosis method for machinery in unsteady operating condition by instantaneous power spectrum and genetic programming. *Mechanical Systems and Signal Processing*, 19(1):175–194, 2005.

[13] L. Zhang, L. B. Jack, and A. K. Nandi. Fault detection using genetic programming. *Mechanical Systems and Signal Processing*, 19(2):271–289, 2005.

[14] S. Silva. GPLAB - a genetic programming toolbox for MATLAB, 2005. http://gplab.sourceforge.net

[15] S. Silva and E. Costa. Dynamic limits for bloat control - variations on size and depth. In K. Deb, R. Poli, W. Banzhaf, *et al.*, editors, *GECCO-2004*, LNCS, 666–677, Seattle, WA, USA, 2004. Springer.

[16] S. Luke and L. Panait. Lexicographic parsimony pressure. In W. B. Langdon, E. Cantú-Paz, K. Mathias, *et al.*, editors, *GECCO-2002*, 829–836, New York, NY, USA, 2002. Morgan Kaufman.