Structured Output Prediction with Hierarchical Loss Functions for Seafloor Imagery Taxonomic Categorization

Navid Nourani-Vatani[1], Roberto López-Sastre[2], and Stefan Williams[3]

[1] ERED, MAN Truck & Bus AG, Munich, Germany
[2] GRAM, University of Alcalá, Spain
[3] ACFR, University of Sydney, 2006 NSW Australia

Abstract. In this paper we study the challenging problem of seafloor imagery taxonomic categorization. Our contribution is threefold. First, we demonstrate that this task can be elegantly translated into a Structured SVM learning framework. Second, we introduce a taxonomic loss function in the structured output classification objective during learning that is shown to improve the performance over other loss functions. And third, we show how the Structured SVM can naturally deal with the problem of learning from data imbalance by scaling the cost of misclassification during the optimization. We present a thorough experimental evaluation using the challenging and publicly available *Tasmania Coral Point Count* dataset, where our models drastically outperform the state-of-the-art-results reported.

Keywords: seafloor imagery, categorization, recognition, structured prediction

1 Introduction

Autonomous Underwater Vehicle (AUV) systems have recently been shown to be effective tools for rapidly and cost-effectively delivering a vast amount of high-resolution, accurately geo-referenced, and precisely targeted optical and acoustic imagery of the seafloor [1]. Processing of this vast amount of collected imagery to label content is difficult, expensive and time consuming. Because of this, typically only a small subset of images are labeled, and only at a small number of points. In order to make full use of the raw data returned from the AUV, this process needs to be automated.

There are, however, many challenges associated with processing images captured underwater. Natural scene illumination may be very poor, and there is often little regular structure with which to delineate objects.

Despite these challenges, there have been many recent advances in the processing of imagery from underwater scenes (e.g. [2,3]). These advances have implications in a diverse range of application areas, including marine ecology, archeology, geology as well as industrial and defense applications.

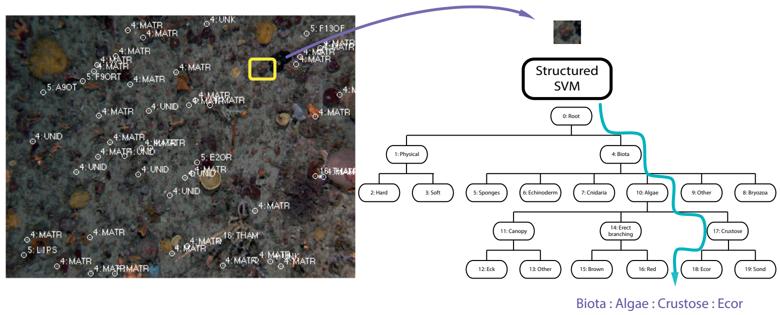


Fig. 1. Overview of our approach. Given a taxonomy, our objective is to classify the image patches according to this. We demonstrate this task can be elegantly translated into the Structured SVM learning framework.

In this paper, we address the task of studying ecosystems and populations from seafloor images. To facilitate this task, we propose an approach to provide marine scientists quantitative data on bottom-dwelling organisms and physical morphology derived from large image archives collected by AUV systems.

For such applications, the state-of-the-art consists of taking a small subset of images, manually labeling the content, and extrapolating to assess distribution and coverage over wider geographical areas, as it has been described in [4]. Essentially, this is a imagery taxonomic categorization problem, see Figure 1, where we are given a pre-determined taxonomy, and the objective is to classify the image patches adhering to this taxonomy.

In this study, we show that this problem can be elegantly translated into a structured learning framework [5], paying special attention to the design of the loss function and potential imbalance in the data set. The main contributions of this paper are: a) We propose a Structured SVM (SSVM) based approach to seafloor imagery classification, and perform a thorough experimental evaluation of a set of taxonomic loss functions; b) We formulate the novel Weighted Hierarchical Difference (WHD) loss, which is able to report the best classification results; c) We show how the Structured SVM can naturally deal with the problem of learning from data imbalance by scaling the cost of misclassification during the optimization; d) We demonstrate that taxonomy-based learning using SSVM yields improved results when hierarchical losses are used, outperforming both standard multi-class SVMs and other hierarchical SVM ensembles [6]; e) A thorough experimental evaluation is reported, using the challenging and publicly available *Tasmania Coral Point Count* dataset [4], where our models drastically outperform the state-of-the-art-results.

The rest of the paper is organized as follows. In the next section, a review of related work, within the context of seafloor image categorization, is given. Section 3 introduces the proposed solution. Section 4 presents the results. Our conclusion is given in Section 5.

2 Related work

There has been substantial research on classification of seafloor species. In [7], starfish detection results from underwater imagery are reported. Approaches for classification of kelp have been also described, e.g. [4]. Multi-class classification has also been attempted, focused mainly on the categorization of different coral species [8,3].

Common for all these approaches is a choice of one or more image-based descriptors and a (collection of) flat classifier(s). More recently, [6] have taken advantage of the taxonomical hierarchy of the species for classification. In a hierarchy with 19 classes, a large framework of 19 binary classifiers, one per node, is employed. The authors presented an in-depth analysis of various training and testing methodologies and have shown state-of-the-art results on their data set.

In this paper we propose the use of an SSVM formulation for the task of taxonomical hierarchical classification. Our approach differs with the approach of [6] in several ways. Firstly, since we are employing a single linear classifier, in the form of an SSVM, it is a much leaner setup with a simplified training and testing strategy. This also signifies, that the amount of training data (and time) is considerably reduced. Secondly, in [6] the hierarchical taxonomy is used *outside* of the classifier through a decision tree. However, we incorporate the taxonomical hierarchy *inside* the loss function of the structured classifier.

Within the same context, i.e. taxonomic categorization, other learning methods have been already proposed to make use of specialist-imposed taxonomies [9,10,11]. Interestingly, in [11], the authors show that the performance of an SSVM [5] based approach can be improved by using an ensemble of local SVMs in some data sets. In our work, we claim that a SSVM with an appropriate hierarchical loss function can efficiently solve the problem, even improving complex ensembles of SVMs [6]. Several hierarchical loss functions are proposed in [9,10], but they differ from the novel Weighted Hierarchical Difference (WHD) loss proposed in this work. In particular, the hierarchical loss in [10] only considers a penalization based on the common ancestor, while our WHD accumulates a loss through the whole hierarchy.

3 Seafloor Imagery Taxonomic Categorization

We formulate the problem of seafloor imagery taxonomic categorization as a structured output prediction problem. We first describe our model, which we propose for solving the taxonomic classification of benthic images using an SSVM [5]. Then, we describe the learning algorithm and introduce the taxonomic loss functions to be evaluated. In order to further improve the performance of the prediction, we propose the novel Weighted Hierarchical Difference (WHD) loss function. Finally, to deal with the problem of learning from data imbalance, we describe a learning strategy, that consists of scaling the misclassified structured predictions during the optimization.

3.1 Model formulation

Using SSVMs, we are able to generalize the SVM to the case of the complex interdependent output space defined by the problem of seafloor imagery taxonomic categorization. Let us assume we are given a collection of N training image patches $\mathcal{I} = \{(x_1, y_1), \ldots, (x_N, y_N)\} \in \mathcal{X} \times \mathcal{Y}$, where $x_i \in \mathbb{R}^d$ encodes the image appearance, and y_i represents the ground truth label of the image in the corresponding taxonomy with a total of C nodes, i.e. $y_i \in \{1, 2, \ldots, C\}$.

With an SSVM we are able to learn a model w associated to a *score* function

$$f(x_i, \hat{y}) =< w, \phi(x_i, \hat{y}) > , \tag{1}$$

which is able to assign a scalar value that indicates how the structured prediction $\hat{y}$ fits the appearance encoded in x_i. Note that during training the objective is to find the classifying hyperplane w for the combined feature representation $\phi(x_i, \hat{y})$ [5].

The specific form of $\phi(,)$ depends on the nature of the problem. Similar to [9], when $\mathcal{Y}$ is taxonomically structured, $\phi(x_i, \hat{y})$ decomposes as $\phi(x_i, y) = \lambda(y) \otimes x_i$, where $\lambda(y)$ is a binary vector that encodes the hierarchical relationship between classes, and $\otimes$ is the Kronecker product, thus $\phi(x_i, y) \in \mathbb{R}^{C \times d}$. In particular, the taxonomy is defined to be an arbitrary lattice (e.g. the tree in Figure 2a), where *all* its elements correspond to categories. It is important to note, that in other taxonomic approaches [9,11] only the minimal elements, i.e. the leaves in the tree, correspond to the categories. In our approach, however, we allow the prediction to be cast at any level of the tree structure. We assume one unique root node, and for every node y in the taxonomy, we define the set of nodes on the path from the root to the node y by $\Omega(y)$. For instance, for the class 10:`Algae`, in the taxonomy shown in Figure 2a, $\Omega(10) = \{0, 4, 10\}$. We then encode this information in a binary vector $\lambda(y)$ for each node y, where the i^{th} element is given by

$$\lambda_i(y) = \begin{cases} 1 \text{ if } i \in \Omega(y) \\ 0 \text{ otherwise.} \end{cases} \tag{2}$$

The corresponding binary vector for class 10:`Algae` in the taxonomy is

$$\lambda(10) = \{1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0\}.$$

3.2 Learning

Our objective is to learn a mapping from sampled features x_i to their corresponding output in the structured output space $\mathcal{Y}$, i.e. the taxonomy. That is, for a set of training images $\mathcal{I} = \{(x_1, y_1), \ldots, (x_N, y_N)\}$, we want to train a *linear* model w that given an image x_i, tends to cast the true structured output for the taxonomy $\hat{y} = y_i$. We formulate this as the following regularized learning problem:

$$\arg\min_{w,\xi} \frac{1}{2} w^\top w + \frac{C}{n} \sum_{i=1}^{n} \xi_i \text{ s.t. } \forall i \,, \xi_i > 0 \,, \tag{3}$$

$$\forall i\,, \forall y \in \mathcal{Y}\backslash y_i :\ \phi(x_i, y_i) - \phi(x_i, y) \geq \Delta(y_i, y) - \xi_i\,, \tag{4}$$

where $\Delta(,)$ is the loss function. The constraint from (4) specifies the following. Consider the i^{th} training image x_i and its corresponding true structured label y_i. We want the true label to score higher than all other hypothesized labellings y.

Intuitively, violating a margin constraint involving a $y \neq y_i$ with high $\Delta(y_i, y)$ should be penalized more severely than a violation involving an output value with smaller loss. This can be accomplished by re-scaling the margin accordingly, as it is shown in Equation (4). The formulation in equations (3) and (4) is often called margin re-scaling [5]. The optimization for the training problem outlined is solved following the cutting plane algorithm in the SVMStruct software package [12].

3.3 Taxonomic Loss Functions

In the formulated structured output taxonomic prediction, different loss functions $\Delta(,)$ can be considered. In this section, we describe all the loss functions that are evaluated in this work. We also introduce the novel Weighted Hierarchical Difference (WHD) loss.

Standard Taxonomic Loss Functions. Given the ground truth label y and the corresponding prediction $\hat{y}$, and considering that $\mathcal{Y}$ is taxonomically structured, we can define the following standard hierarchical loss functions.

We consider three distinct loss functions: 1) the distance to the nearest ancestor in the tree $\Delta_n(\hat{y}, y)$; 2) the classical distance through the tree $\Delta_t(\hat{y}, y)$ [13]; 3) and the hamming distance $\Delta_h(\hat{y}, y) = \sum_i |\lambda_i(\hat{y}) - \lambda_i(y)|$ [11], which counts the number of non-shared nodes on the path between the true class y and the prediction $\hat{y}$

We also consider it important to analyze the performance of taxonomic loss functions, that explicitly incorporate the hierarchical statistics for true positives (tp), false positives (fp), false negatives (fn) and true negatives (tn). These can be efficiently obtained from the taxonomy and with y and $\hat{y}$. Once these hierarchical statistics are computed, we proceed to define the following taxonomic loss functions: $\Delta_{precision}(\hat{y}, y) = tp/(tp+fp)$, $\Delta_{recall}(\hat{y}, y) = tp/(tp+fn)$, $\Delta_{accuracy}(\hat{y}, y) = tp/(tp+fp+fn)$, $\Delta_{hier.hamming}(\hat{y}, y) = (tp+tn)/(tp+fp+fn+tn)$ and $\Delta_{f_1}(\hat{y}, y) = (2 \cdot tp)/(2 \cdot tp + fp + fn)$.

Finally, one simple loss, which can be incorporated to our approach, is the standard 0/1 loss function

$$\Delta_0(\hat{y}, y) = \begin{cases} 0 \text{ if } \hat{y} = y \\ 1 \text{ otherwise.} \end{cases} \tag{5}$$

Note that $\Delta_0(\hat{y}, y)$ transforms our taxonomic classification problem into a standard multi-class prediction problem (with each node of the taxonomy representing a different, and unrelated, class).

Weighted Hierarchical Difference. One of the main limitations of the hamming distance $\Delta_h(\hat{y}, y)$ [11], is that it does not penalize an error higher up the hierarchy more severely; an aspect we consider fundamental for taxonomic categorization problems. Consider a classifier mistaking a Yellow Labrador (*Canis lupus familiaris*) to a Persian Cat (*Felis catus*) or mistaking the Yellow Labrador with a Golden Retriever. The former is obviously a bigger mistake and should incur a greater penalty.

To highlight this difference, in this paper, we propose the following loss, which we call the Weighted Hierarchical Difference (WHD) loss,

$$\Delta_{\mathrm{WHD}}(\hat{y}, y) = \sum_i \left| \Psi(\lambda_i(\hat{y})) - \Psi(\lambda_i(y)) \right|. \tag{6}$$

Essentially, this WHD computes the L^1-norm of the difference of vectors $\Psi(\lambda(\hat{y}))$ and $\Psi(\lambda(y))$. We define $\Psi()$ as a weighting function, which divides each component i of the binary vector $\lambda(y)$ by the level it belongs to. For instance, in the taxonomy shown in Figure 2a, for the class 10:`Algae`, we have that $\Psi(\lambda(10)) = (1, 0, 0, 0, 1/2, 0, 0, 0, 0, 0, 1/3, 0, 0, 0, 0, 0, 0, 0, 0, 0)$.

We have incorporated all these loss functions to the SVMStruct package [12]. A reference implementation of the code has been made publicly available[4].

3.4 Dealing with imbalanced taxonomies

Data imbalance is a typical problem of taxonomic data sets. The larger the number of classes in the hierarchy, the more difficult it is to guarantee that all classes are assigned a similar number of samples when collecting the data. In this section, we show that learning from highly imbalanced data is a problem that can be naturally addressed employing a weighting strategy for the cost of a misclassification during learning of the SSVM.

Learning from imbalanced data is problematic. As it is shown in [14], an SVM learned with an imbalanced data set can be skewed and become unfavorable to the minority class. Different techniques have been proposed to deal with this problem.

Oversampling is a data preprocessing technique, that balances the data set before training [15]. Basically, the minority classes are oversampled in order to get a data set where all classes have similar number of samples. However, both the training time and the memory requirements of the algorithm naturally increase. Furthermore, if non-linear kernels are used, the test time might increase too.

In order to solve these oversampling problems, we formulate an SSVM approach where, we propose to dynamically weight the taxonomic loss function during the optimization according to the number of samples per class. This weighting is chosen such that a misclassification of a small class is penalized more.

We proceed to define different error costs for the different classes in our taxonomy. In a standard binary SVM classifier, this is accomplished by adjusting

[4] `https://github.com/nourani/Seafloor_SSVM`

the cost parameter to $C^+ = \alpha^+ \times C$ and $C^- = \alpha^- \times C$, where C is the original cost parameter and α^+/α^- are weight constants for the large and small classes, respectively. These weight constants are often set to the inverse of the class size ratios.

In our SSVM formulation, the cost of a misclassification can be adjusted by adding a weight penalty α_i to the loss, i.e. $\Delta(y_i, y)/\alpha_i$. So, (4) is transformed as follows,

$$\forall i\,, \forall y \in \mathcal{Y} \backslash y_i : \; \phi(x_i, y_i) - \phi(x_i, y) \geq (\Delta(y_i, y)/\alpha_i) - \xi_i\,. \tag{7}$$

We define α_i as the penalty for an incorrect classification of label y_i. In our experiments we follow the standard procedure of setting the penalty to the inverse of the class instance ratios: $\alpha_i = n_i/N$, where n_i is the number of instances in class i and N is the total number of training instances in the taxonomy. For small classes, α_i will be small and the loss for misclassification will become larger, compensating for the lack in training instances.

The experimental validation shows that our weighting strategy outperforms the oversampling method, both in classification accuracy and runtime.

4 Results

4.1 Experimental Setup

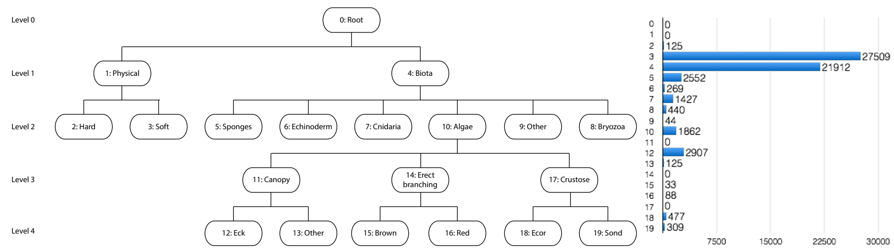
For our experimental validation, we use the publicly available *Tasmania Coral Point Count* data set, introduced in [4], which consists of 1258 images labeled by expert marine scientists, with each image containing 50 labels at 50 randomly selected pixels, adhering to the taxonomical hierarchy shown in Figure 2(a). There are more than 130 different species labels in the data set, which have been collapsed to 19 classes (c.f. Figure 2(b)) with the guidance of marine scientists.

From Figure 2(b) it can be seen that the data set is extremely unbalanced. Some of the parent nodes contain no instances, while other parent nodes do contain labels. What this signifies is that not all labels go to the leaf nodes. It is necessary to formulate a classification strategy, that both takes into account the unevenness and the labeling philosophy of the data set.

For all the experiments, the data set is divided into 80% training samples and 20% test samples following the experimental setup in [4]. Exponential grid search in the range $10^{-2} - 10^{2}$ is used to find the cost parameter, C, for the SSVM classifier proposed, by training on 2/3 of the training samples and validating on the last 1/3. For the visual features, we follow [6] and for each image patch (of 31×31 pixels) we compute the Histogram Fourier Local Binary Patterns [16] (LBP-HF) descriptor. We evaluate the performance of our solutions using the modified hierarchical F1-score introduced in [6].

4.2 Seafloor Imagery Taxonomic Categorization

We start the experimental validation assessing the performance of the different taxonomic loss functions proposed. We compare our results with the baseline



(a) (b)

Fig. 2. a) The hierarchy, and b) number of instances per class in the *Tasmania Coral Point Count* data set.

Table 1. Performance of the 10 loss functions on the *Tasmania* data set.

Loss	H-F1$_{mod}$	Accuracy	Precision	Recall
Δ_0	88.67	**67.58**	16.41	17.99
Δ_{WHD}	**89.06**	**67.58**	17.19	**22.66**
Δ_{f_1}	**89.06**	**67.58**	16.40	15.23
$\Delta_{precision}$	88.28	66.02	13.28	15.63
Δ_{recall}	87.50	62.89	17.58	16.02
$\Delta_{accuracy}$	**89.06**	**67.58**	16.79	17.19
$\Delta_{hier.ham}$	**89.06**	67.19	16.02	22.27
Δ_n	77.73	50.00	**18.36**	17.97
Δ_t	84.38	67.19	15.63	15.63
Δ_h	84.38	**67.58**	15.23	15.63

Table 2. Comparison with a flat multi-class SVM classifier scheme and the state-of-the-art.

Method	H-F1$_{mod}$ %
Flat multi-class	66.67
MPS$_{0.5}$ [6]	80.24
Δ_{WHD} (Ours)	**89.06**

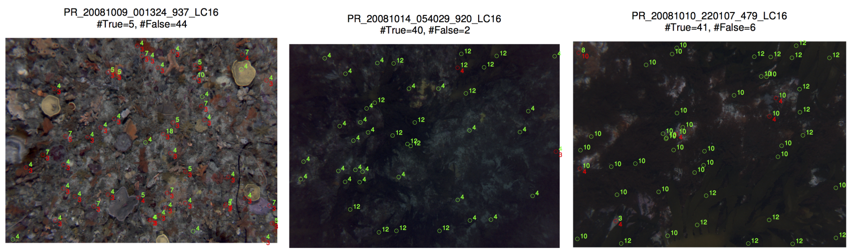
methodology, which consists of a multi-class SSVM [5] employing a 0/1-loss, i.e. Δ_0. We have used the code available for multi-class classification at [12].

Table 1 shows the performances of the loss functions in terms of the hierarchical F1-score, accuracy, precision and recall. We observe that the new WHD loss function is performing the best. We also compare our method to a flat multi-class SVM classifier and the state-of-the-art reported on the same data set by [6] (see Table 2). The authors of [6] presented a one-classifier-per-node hierarchical approach employing max probability switching with thresholding, MPS$_{0.5}$. For the flat SVM classification scheme, we have trained a standard multi-class linear SVM, using the LibSVM [17] package. Table 2 shows that our SSVM approach significantly improves both the baseline and the state-of-the-art MPS$_{0.5}$.

The results reported illustrate some important outcomes. Firstly, employment of SSVM has greatly improved the performance over the one-node-per-classifier approach of [6] from 80.24% to 89.06%. Secondly, this large improve-

Table 3. Main results analyzing the data imbalance on the sampled test set.

Method	Prec.	Rec.	H-F1$_{mod}$	Samples [n]	Time [s]	Mem [GB]
Std. SSVM	16.41	7.42	69.53	48,181	544	**0.5**
Std. SSVM + Oversampling	15.39	**13.00**	72.27	305,474	3259	5.2
Weighted SSVM	**18.48**	9.13	**73.05**	48,181	**228**	**0.5**
Weighted flat SVM (LibSVM)	7.27	3.30	51.32	48,181	11887	1.0



a b c

Fig. 3. Qualitative results. The location of the ground truth label is shown with a $\circ$ marker and green text. Red marker indicates incorrect classification. a) One of the worst results on the test set. b-c) Two of the best classification results.

ment of 8% has come in addition to a considerably simplified methodology and setup. We can conclude that taxonomy-based learning using SSVM yields improved results when hierarchical losses are used, outperforming both standard multi-class SVMs and other hierarchical SVM ensembles [6]. Finally, Figure 3 shows some qualitative results.

4.3 Learning from imbalanced data

Here we propose an experimental validation to analyze the influence of the data imbalance on the described solutions. We start evaluating the performance of a standard SSVM with and without oversampling. We also evaluate the proposed approach, described in Section 3.4, named Weighted SSVM. For the sake of comparison, we also report the performance of a weighted flat multi-class SVM. Note that, to ensure a proper evaluation, we oversample the test data and draw an equal number (847) of instances for each class, while maintaining the overall total number of test samples (11854). When an SSVM is used, we report the performance obtained using our novel WHD loss function.

Table 3 shows the main results. All times reported are on a 2.3GHz i7 processor. We observe that the proposed Weighted SSVM obtains the best results in terms of the H-F1 score and precision. This has been accomplished by increasing both the precision and the recall by $\sim 2\%$ compared to the Std. SSVM. Our Weighted SSVM also finishes the training in almost half the time of the Std. SSVM. The SSVM with oversampling does not outperform our Weighted SVM even though it uses $5\times$ as much training data, resulting in a ten-fold increase in both processing time and memory usage compared to the proposed Weighted SSVM. Our weighted approach is clearly the best approach for dealing with imbalance in the training data.

5 Conclusion

A novel approach to seafloor imagery taxonomic categorization has been developed. We have evaluated the incorporation of the taxonomy inside the loss function of an SSVM formulation. We have also introduced the novel WHD loss function, whose results show a significantly better performance compared to the state-of-the-art on a new challenging underwater data set. We have further demonstrated that it is possible to follow a weighting strategy in the SSVM optimization to alleviate imbalance in the training data set.

Acknowledgements. The authors acknowledge the Australian National Research Program (NERP) Marine Biodiversity Hub for the taxonomical labeling and the Australian Centre for Field Robotics for gathering the image data.

References

1. Williams, S., Pizarro, O., Jakuba, M., Johnson, C., Barrett, N., Babcock, R., Kendrick, G., Steinberg, P., Heyward, A., Doherty, P., et al.: Monitoring of benthic reference sites: Using an autonomous underwater vehicle. IEEE Robotics & Automation Magazine **19**(1) (2012) 73–84
2. Johnson-Roberson, M., Kumar, S., Williams, S.: Segmentation and classification of coral for oceanographic surveys: A semi-supervised machine learning approach. In: IEEE OCEANS. (2006) 1–6
3. Beijbom, O., Edmunds, P.J., Kline, D., Mitchell, B., Kriegman, D.: Automated annotation of coral reef survey images. In: IEEE CVPR. (2012) 1170–1177
4. Bewley, M., Douillard, B., Nourani-Vatani, N., Friedman, A., Pizarro, O., Williams, S.: Automated species detection: An experimental approach to kelp detection from sea-floor auv images. In: ACRA. (2012)
5. Tsochantaridis, I., Hofmann, T., Joachims, T., Altun, Y.: Support vector machine learning for interdependent and structured output spaces. In: ICML. (2004) 104
6. Bewley, M., Nourani-Vatani, N., Rao, D., Douillard, B., Pizarro, O., Williams, S.: Hierarchical Classification in AUV Imagery. In: Field and Service Robotics. (2013)
7. Smith, D., Dunbabin, M.: Automated counting of the northern pacific sea star in the derwent using shape recognition. DICTA (September 2007) 500–507
8. Soriano, M., Marcos, S., Saloma, C., Quibilan, M., Alino, P.: Image classification of coral reef components from underwater color video. In: MTS/IEEE Conference and Exhibition OCEANS. (2001) 1008–1013
9. Cai, L., Hofmann, T.: Hierarchical document categorization with support vector machines. In: ACM CIKM. (2004) 78–87
10. Tuia, D., Muñoz Marí, J., Kanevski, M., Camps-Valls, G.: Structured output svm for remote sensing image classification. Journal of Signal Processing Systems **65**(3) (2011) 301–310
11. Binder, A., Müller, K., Kawanabe, M.: On taxonomies for multi-class image categorization. International Journal of Computer Vision (IJCV) **99** (2012) 281–301
12. Joachims, T.: Multi-class support vector machine. `http://www.cs.cornell.edu/people/tj/svm_light/svm_multiclass.html` (2008)
13. Wang, K., Zhou, S., Liew, S.: Building hierarchical classifiers using class proximity. In: International Conference on Very Large Data Bases (VLDB). (1999)
14. Veropoulos, K., Campbell, C., Cristianini, N.: Controlling the sensitivity of support vector machines. In: IJCAI. Volume 1999. (1999) 55–60
15. Japkowicz, N., Stephen, S.: The class imbalance problem: A systematic study. Intelligent data analysis **6**(5) (2002) 429–449
16. Ahonen, T., Matas, J., He, C., Pietikäinen, M.: Rotation invariant image description with local binary pattern histogram fourier features. Image Analysis (2009) 61–70
17. Chang, C., Lin, C.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology **2** (2011) 27:1–27:27