# Automated Annotation of Coral Reef Survey Images

Oscar Beijbom[†] Peter J. Edmunds[*] David I. Kline[‡] B. Greg Mitchell[‡] David Kriegman[†]
{obeijbom, dkline, gmitchell, kriegman}@ucsd.edu, peter.edmunds@csun.edu
{[†]Department of Computer Science and Engineering, [‡]Scripps Institution of Oceanography},
University of California, San Diego. [*]Department of Biology, California State University Northridge.

## Abstract

*With the proliferation of digital cameras and automatic acquisition systems, scientists can acquire vast numbers of images for quantitative analysis. However, much image analysis is conducted manually, which is both time consuming and prone to error. As a result, valuable scientific data from many domains sit dormant in image libraries awaiting annotation. This work addresses one such domain: coral reef coverage estimation. In this setting, the goal, as defined by coral reef ecologists, is to determine the percentage of the reef surface covered by rock, sand, algae, and corals; it is often desirable to resolve these taxa at the genus level or below. This is challenging since the data exhibit significant within class variation, the borders between classes are complex, and the viewpoints and image quality vary. We introduce Moorea Labeled Corals, a large multi-year dataset with 400,000 expert annotations, to the computer vision community, and argue that this type of ecological data provides an excellent opportunity for performance benchmarking. We also propose a novel algorithm using texture and color descriptors over multiple scales that outperforms commonly used techniques from the texture classification literature. We show that the proposed algorithm accurately estimates coral coverage across locations and years, thereby taking a significant step towards reliable automated coral reef image annotation.*

## 1. Introduction

In many scientific disciplines experts routinely analyze large quantities of image data. However, not only has the capacity for acquiring digital images vastly outpaced the resources to manually annotate those images, but there are also issues with lack of consistency and objectivity in human labeling [18, 8]. One such domain is coral reef ecology, which is particularly important given the crucial ecological roles of coral reefs, and their current state of decline in health and abundance [26]. To understand this development, ecologists need accurate and large-scale coral reef coverage data. As satellite images are ineffective for this purpose [14] and low altitude photography suffers from problems such as surface effects, coral ecologists commonly do *in situ* studies. Recent innovations in image acquisition techniques, such as autonomous underwater vehicles [24] and towed diver sleds [15] have greatly

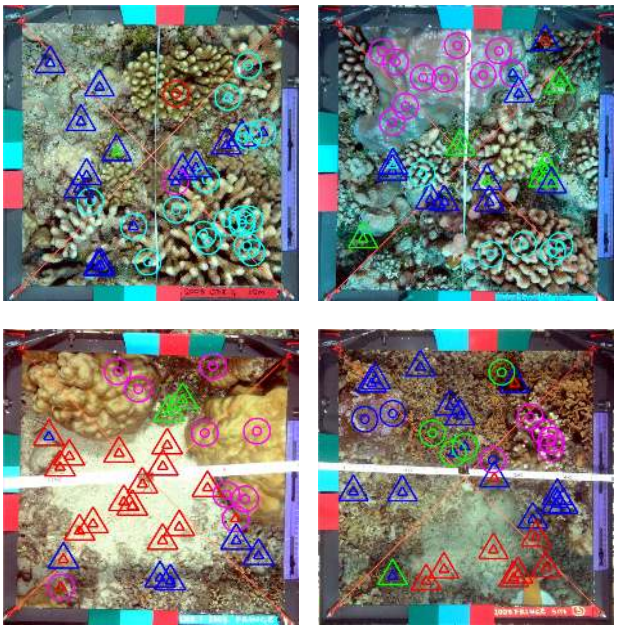


Figure 1. **Moorea Labeled Corals example images:** Top row images are from the *Outer 10m reef* habitat, bottom from the *fringing reef*. Superimposed on each image is a subset of the ground truth annotations (smaller symbols) and the estimated classifications by the proposed algorithm (larger symbols). Circles represent coral genera *Acropora*, *Pavona*, *Montipora*, *Pocillopora*, *Porites* and triangles are non-coral substrates, Crustose Coralline Algae, Turf algae, Macroalgae and Sand. Note the organic borders between the substrates and complex class morphologies. The white transect line going through the images is part of the sampling methodology, as is the metal frame seen along the edges. This figure is best viewed in color.

increased the number of images available and researchers, reef managers, and government agencies such as the National Oceanic and Atmospheric Administration (NOAA), routinely collect millions of images per year as part of reef surveys. However, the lack of automated annotation methods effectively caps the monitoring effort since manual annotation is slow, tedious, and expert time is limited; anecdotally, as little as 1-2 % of captured images gets annotated.

Concurrently, there have been great advances in recognition of specific objects, object classes, and scenes on widely used benchmark data sets (*e.g.*, Caltech 101 [11], Pascal [10]), and there is intense competition within the research community to develop new methods. Yet recognition in much of the biological image data introduces new challenges. As noted by Adelson [1], rather than trying to recognize *things* as found in most of the object recognition datasets, biological datasets often require identifying *stuff*. In the case of coral reef image analysis the size, color, shape, and texture of each of the classes may vary significantly and the boundaries are often organic and ambiguous. Further, water turbidity can change dramatically between sites and years due to currents, plankton, algal blooms, etc., affecting ambient light and image colors. We argue that this type of data pose challenges quite different from those posed by the above mentioned datasets, and also from benchmarks commonly used for texture based recognition, Brodatz [25] and CUReT [9].

Due to the above mentioned challenges, standard annotation techniques such as image labels, bounding boxes or full segmentations are inappropriate. A full manual segmentation would be too time consuming to collect and neither bounding boxes nor gross image labels provide the level of detail needed. Instead, coral ecologists commonly rely on random point sampling. In this process images are annotated by asking an expert to identify at a specific point (pixel) in the image, which class is present (most prevalent) at that single location. The process is repeated for a fixed number of randomly chosen locations in the image (typically 10-200) and often facilitated by software such as Coral Point Count [16]. Once annotated, coverage statistics can be calculated over sets of images. A typical survey might include 1000 images with 200,000 annotations requiring >400 person hours to label. These expert annotated images are a bonanza for computer vision researchers – a rich source of labeled and relevant data.

**Related work:** In a recent effort to automate coral classification, Stokes and Deane [28] devised a method that utilizes the discrete cosine transform and a k-nearest neighbor classifier to do patch-wise classification of benthic substrates. Their work shows promising results, but their small dataset (16 images) makes it hard to draw conclusions on scalability. Other work includes [21, 27], that use texture and color descriptors, but presume the availability of patches hand cropped from the image data.

There is a rich literature on object recognition and texture-based classification in the computer vision community. Dana *et al*. created the CUReT database in 1999, and this became a popular benchmark for texture recognition [9]. Malik *et al*. [19] made an important contribution by introducing the texton as a cluster center of filter responses, and later Cula and Dana [7], were able to do robust classification from a single image. Varma and Zisserman further improved the CUReT benchmarks by introducing the rotational invariant Maximum Response (MR) filter bank and reported classification accuracies of up to 98.61 % [31]. Some years later Varma and Zisserman proposed the use of patch exemplars instead of filter responses [32]. They report a minor improvement on the CUReT database, but in general the classification scores seem to have saturated in the high nineties. One of the limitations of CUReT is that all objects are cropped and each image contains only one type of texture. Winn *et al*. [33] introduced the MSRC dataset in 2005 as an alternative benchmark for texture based methods, but this suffers from other problems as pointed out by Torralba and Efros [29]. In contrast to CURet, MSRC, and most other datasets used for performance benchmarks, the Moorea Labeled Corals (MLC) dataset was *not* created explicitly for the purpose of developing computer vision methods. We believe this provides a better opportunity for benchmarking, and indicates a direction for future computer vision challenges.

**Contributions:** The contribution of this paper is twofold: (1) We introduce the analysis of coral reef survey images as a relevant and important application domain to the computer vision community and propose a benchmark dataset. (2) We propose a method based on texture and color descriptors over multiple scales that outperforms standard texture classification methods and establishes a strong baseline on the MLC dataset. This is, to the best of our knowledge, the first paper to address automated annotations of coral reef survey image on a large scale.

## 2. Dataset

The Moorea Coral Reef Long Term Ecological Research (MCR LTER) project has been collecting image data from the island of Moorea (French Polynesia) since 2005. This project monitors six sites around the island, and four habitats at each site. Each year, transect lines are attached between points permanently anchored in the reef. Then, divers move along the transect line, acquiring images using a digital SLR equipped with underwater strobes. The camera is attached to a frame (50 cm by 50 cm) to keep the distance from the bottom constant and the camera orientation parallel to the reef surface. Sample images are provided in Fig. 1 where the frame is seen along the image edges, and the white transect lines pass through the middle. The orange

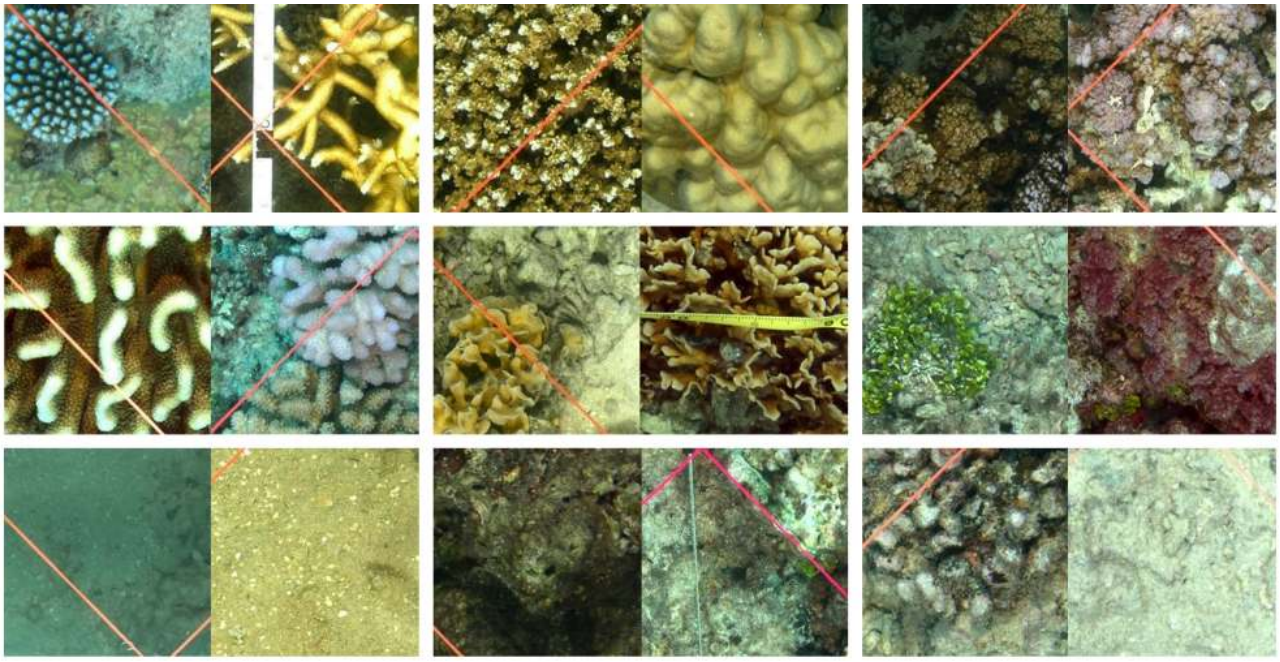


Figure 2. **Moorea Labeled Corals classes:** Each class is represented by two example patches. First row: *Acropora, Porites, Montipora*. Second row: *Pocillopora, Pavona*, Macroalgae. Third row: Sand, Turf algae, CCA. Each class exhibits great variability with respect to growth form, color, scale and viewing angle. For example, note the two different growth forms of *Acropora* and *Porites*, and the difference in color and scale of the *Pocillopora* patches. Macroalgae varies tremendously in shape and color, and often projects from underneath the corals. Both CCA and Turf algae tend to overgrow dead coral, which pose a challenge for automated analysis since the coral skeleton retains its shape, but now classifies as the the algae that overgrows it. Also, CCA and Turf are similar and are hard to discriminate from an RGB photography. Finally, we point out that the images in this grid has been hand picked to clearly exemplify each class. A random selection of patches would include a large amount of patches that straddles class boundaries, and where class designation is less clear.

lines are attached to the frame to help the diver align the center of the frame with the target. Even with this careful method of acquisition there are large differences in image appearance as mentioned in Sec. 1. Across years things become even more challenging as the ecosystem changes. On this reef an outbreak of Crown of Thorns Sea Star (a predatory sea star) in 2006, followed by Cyclone Oli in February 2010, together destroyed the majority of the corals. Data from 2011 show signs of rapid regrowth of the reef through recruitment. All images are labeled using the random point sampling method mentioned in Sec. 1. In this dataset 200 points were laid over on each image at random locations and then labeled by an expert according to an established taxonomy ($\sim$ 30 categories) for the site.

**Moorea Labeled Corals:** The MLC dataset is a subset of the MCR LTER packaged for computer vision research. It contains 2055 images from three habitats: *fringing reef*, *outer 10m* and *outer 17m*, from 2008, 2009 and 2010. It also contains random point annotation (row, col, label) for the nine most abundant labels, four non coral labels: (1) Crustose Coralline Algae (CCA), (2) Turf algae, (3) Macroalgae and (4) Sand, and five coral genera: (5) *Acropora*, (6) *Pavona*, (7) *Montipora*, (8) *Pocillopora*, and (9) *Porites*. These nine classes account for 96% of the annotations and total to almost 400,000 points. There is a large variation in the number of samples from each class, Fig. 5. The MLC dataset is available at `http://vision.ucsd.edu/data`.

**Labeling ambiguity:** The annotations in this settings are not 'ground truth' in the common sense, meaning an absolute truth. Culverhouse *et al*. studied this phenomena for plankton classification and reported surprisingly weak inter and intra operator consistency rates [8]. The authors list several reasons for this: short term human memory, fatigue and boredom, recency effects (bias towards labels recently used) and positivity bias (bias towards what one expects to see) [8]. Further, the labels may be noisy for reasons outside the control of the human operator. For example, in the present case, there might be insufficient signal in an RGB image to discriminate Turf algae and CCA. Ninio *et al*. investigated some of these effects for underwater video images of the Great Barrier Reef. They report accuracies of 96% for hard corals, 92.5% for soft corals and 80.6% for algae when compared to *in situ* observations. They also report on inter- and intra-operator variabilities but only on aggregated coverage statistics [23]. Carleton & Done ob-

served a "far inferior" ability to discriminate among corals using video transects compared to *in situ* observations [3]. In general there is an acknowledgement within the community that errors exists in distinguishing among benthic groups and that it can be severe in a few cases, for example discerning between Turf algae and CCA. This notion is supported by a recent (unpublished) NOAA study comparing the annotations of a benthic expert to those of moderately trained operators. Their data indicate 97% agreement between the operators and the expert on the binary task of discriminating between corals and non-corals, 85% agreement on the task of discriminating between the 5 most abundant hard corals, and 79% agreement when discriminating between Macroalgae, Turf algae, CCA, and Sand [12]. We are currently studying these effects for four Pacific datasets including the MLC dataset.

## 3. Method

In this section we describe the proposed method. As the objects we aim to classify are small, lack clear boundaries and a clear sense of shape, we represent them using texture descriptors. Varma and Zisserman [31] proposed a method where textures are modeled as histograms of textons [17], and introduce a rotational invariant filter bank, Maximum Response (MR). Their method is straight forward and achieves accuracy of >95% on the CUReT database [9]. Driven by difficulties in the problem domain, the proposed method extends [31] in two ways: the use of multiple scales and the incorporation of color information. As shown in Fig. 3, and further discussed in Sec. 3.3, there is no straightforward way to choose a fixed patch size over which to integrate image information, which led us to a data-driven, multiple patch approach. The color extension was introduced to encode important color information. However, this is not trivial as colors are corrupted under water, *e.g.*, the red colors are attenuated as a function of depth [4]. We discuss this further in Sec. 3.2.

### 3.1. Preprocessing

All images are first downsized by a factor of 2. We then apply a method, *ColorChannelStretch*, which, for each color channel individually, finds the 1% and 99% intensity percentile, subtracts the lower from all intensities in that channel (and sets any negative pixel values to 0), and then divides by the upper. This method slightly modifies the image colors, and was found to be empirically superior to *IntensityStretch* which stretches the global image intensities across all channels, Table 1.

### 3.2. Texture and Color

We use the Maximum Response (MR) filter bank introduced in [31] which encodes rotational invariance by first filtering with bar and edge filters at different orientations and then outputting the maximum over the orientations. It also contains a circular Gaussian and Laplacian filter. By cross validating over different sizes we arrive at bar and edge filters with standard deviations of 1, 3 and 8 pixels along the short dimension, and circular filters standard deviation of 6 pixels, thus producing an 8 dimensional filter output vector. The filter sizes correspond to 0.5, 1.5, 4 and 3 mm actual distance on the reef surface which is small enough to capture the polyps characterizing the different coral types, along with the finer coral structures. After filtering, following [20, 22], we apply a contrast normalization,

$$\mathbf{F}(\mathbf{x}) \leftarrow \mathbf{F}(\mathbf{x})[\log(1 + L(x)/0.03)]/L(x) \qquad (1)$$

where $L(x) = ||\mathbf{F}(\mathbf{x})||_2$ is the magnitude of the filter response vector, $\mathbf{F}(\mathbf{x})$ at pixel $x$.

**Color:** The method as presented in [31] does not give special care to color. While colors are not generally a robust source of information [2], particularly underwater with artificial light sources [30], we show that color information is useful for this task. We encode color information by applying the filters to each color channel in the *L*a*b** color space and then stack the filter response vectors. Ideally, when using color descriptors for discrimination, color correction should be performed. We were however unable to do this as: (1) there was no reliable color references in the images, (2) there is a deficiency in the literature of methods for underwater color correction. This issue was left to

| | L*a*b* | HSV | RGB | Gray |
|---|---|---|---|---|
| **Intensity - Stretch** | 74.6% | 65.3% | 71.8% | 69.5% |
| | 63.9% | 58.9% | 64.3% | 63.6% |
| | 80.1% | 79.1% | 78.2% | 76.4% |
| | [72.9%] | [67.8%] | [71.4%] | [69.8%] |
| **Color - Channel - Stretch** | 74.7% | 72.6% | 72.5% | 70.1% |
| | **67.4%** | 64.9% | 66.0% | 64.5% |
| | 80.1% | 80.6% | 80.8% | 79.8% |
| | **[74.1%]** | [72.2%] | [73.1%] | [71.1%] |
| **None** | **75.1%** | 73.2% | 72.7% | 70.4% |
| | 64.1% | 59.3% | 65.8% | 63.9% |
| | 81.3% | **82.3%** | 80.8% | 79.2% |
| | [73.5%] | [71.6%] | [73.1%] | [71.2%] |

Table 1. **Color space and preprocessing method:** The percent of correctly predicted points; in each cell the first row represents the 2008 $\Rightarrow$ 2008, experiment, the second 2008 $\Rightarrow$ 2009, the third 2008 + 2009 $\Rightarrow$ 2010, and the last the average across all experiment (see Sec. 4). The best score for each experiment is represented in bold, and the lower right cell represent base-line performance. While the differences are subtle and noisy, we see that the L*a*b* color space and *ColorChannelStretch* works best, in particular for the across year experiments. By stretching the image histogram we compensate for differences in water turbidity and lighting. The comparison shows that color information is useful even though absolute color references are unavailable.

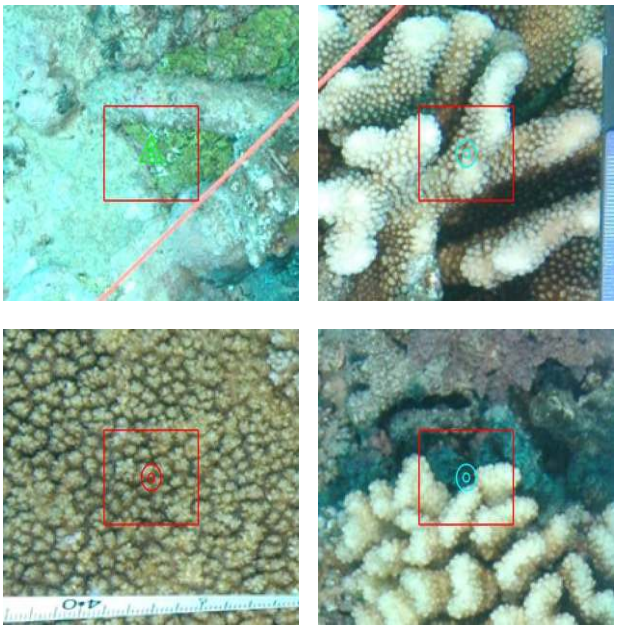


Figure 3. **Patch size problems:** Four different patches from the dataset, each with a region of 61 by 61 pixels indicated with a red box. In top left, the signal from the small algal patch might be overwhelmed by the surroundings if the patch is too big. In top right, on the other hand, a small patch would not capture the larger textures. Also, in bottom left, the histogram of visual words would be more stable if a larger area is used. Bottom right show a case where the point of interest is on the edge between classes, in which case any patch size will be problematic.

future work, and we settle for intensity normalization by histogram stretching as discussed in Sec. 3.1. An empirical comparison on the end-to-end system shows that *ColorChannelStretch* and *L*a*b** yield the most consistent results across years (Table 1). It seems that by stretching the image histogram we are able to compensate for some differences in water turbidity and lighting.

### 3.3. Descriptors

To create the texton maps, a dictionary of textons is required. For this purpose we select a subset of the 2008 images that were not used in any test sets. Filter responses from each of the nine classes were separately aggregated across the images, and k-means [13] clustering with 15 cluster centers was applied to each set of filter responses. Finally the cluster centers, or *textons*, from the different classes were merged to create the dictionary with 135 words, each 24 dimensions.

To extract the texture descriptor, we first apply the filters over the whole image. This yields a 24-dimensional feature vector for each pixel in the image. The filter responses are then mapped to the texton with smallest 2-norm distance, creating an integer valued textonmap. The actual feature

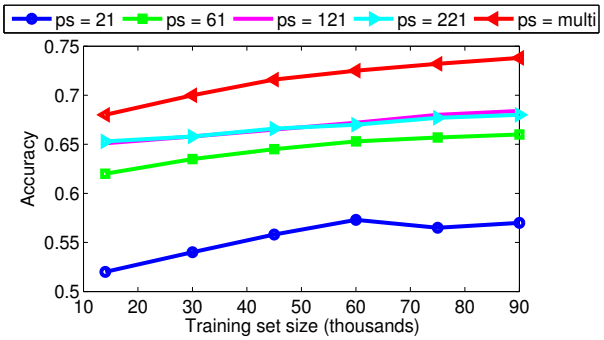



Figure 4. **Patch size experiment:** End to end performance result using different patch sizes (ps). The $ps = multi$ results are generated using all four integration areas as described in Sec. 3.3. For size reference, note that the patches in Fig. 2 are all 221 by 221 pixels. All runs are evaluated on the same test set comprising 200 images from 2008. The training data comprises subsets of the remaining images from 2008. Note how the single patch size performance levels out at $ps = 61$. From this experiment it is evident that by concatenating the histogram counts over multiple scales, a more discriminative descriptor is generated.

vector, or descriptor, is set to be the normalized histogram of textons around a patch of interest. As illustrated in Fig. 3, the selection of an appropriate patch size is not trivial. We evaluate the performance across four different patch sizes $(21, 61, 121$ and $221)$ pixels, and also using a concatenation of the descriptors (bin counts) across all scales. The evaluation shows that the multi-scale approach is superior to any of the single patch sizes. This is illustrated in Fig. 4 where we sweep over training set sizes. The complete image processing pipeline is summarized in Algorithm 1 where $I$ is the input image, $MR$ the filter bank, $D$ the dictionary of textons, $A$ the row and column index of the annotated image points (with a total of $n$ points), and $T$ the texton map.

### 3.4. Machine Learning

The LIBSVM implementation of Support Vector Machines (SVM), with Radial Basis Function kernel were used throughout this work [6, 5]. For each experiment the RBF-SVM training step is preceded by 4-fold cross validation on the training data, where parameters $\gamma$ and $C$ are optimized by a logarithmic grid search over values $\log(\gamma, C) \in \{-5, -4, \ldots, 4, 5\}$. The training data was subsampled so that a maximum of 15000 samples from each class is represented; the subsampled classes get assigned a weight inversely proportional to the subsample rate.

## 4. Experiments

We evaluate end-to-end system performance by running three experiments. In 2008 $\Rightarrow$ 2008, we trained on two thirds of the images from 2008 and tested on the last third. In 2008 $\Rightarrow$ 2009 and 2008 $+$ 2009 $\Rightarrow$ 2010 we evalu-

**Algorithm 1** Extract Descriptors(I, MR, D, A)

$I \leftarrow \text{COLORCHANNELSTRETCH}(I)$
$I \leftarrow \text{CONVERTTOLAB}(I)$
$F \leftarrow \text{APPLYFILTERS}(I, MR)$
**for** each pixel, $p \in I$ **do**
  $T_p = \arg\min_w ||F_p - D_w||_2$
**end for**
**for** $i = 1$ to $n$ **do**
  $D_i \leftarrow \text{MULTISCALEHISTOGRAMS}(T, A_n)$
**end for**

ate across-year performance. The last two experiments are of particular interest for this application as they explore the opportunity to let the machine annotate all data from a certain year. They are also challenging since the appearance of the reef changes between years as discussed in Sec. 2. The training and testing data are created by the procedure described in Algorithm 1, and then run through the SVM. Precision is determined as the ratio of correctly classified patches in relation to the expert annotation. We also considered using the f-score, or harmonic mean which is often used for multi class classification performance evaluation, but this was too noisy due to the small number of samples in some of the classes. In addition we evaluate performance by comparing coverage statistics extracted from the estimated annotations to those extracted from the expert annotations. Percent cover is calculated as the ratio of annotations of a certain type compared to the total number of annotations.

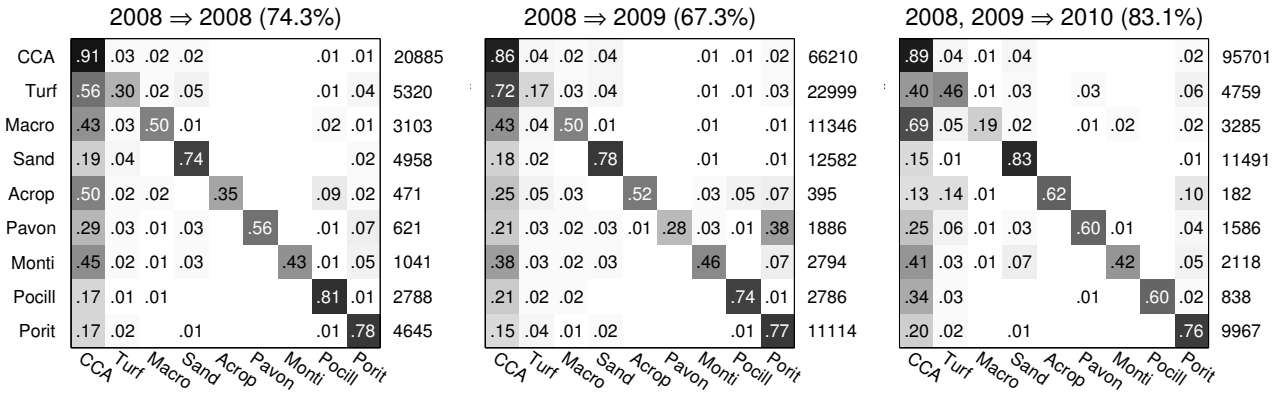
**Patchwise results:** Table 2 shows results for the full experiment, and three important sub classification tasks. It also shows random classification (based on the prior probabilities) scores as these get quite large in some cases due to the abundance of CCA. Best performance on the full labelset is achieved on the within year experiment (74%). For the second experiment, $2008 \Rightarrow 2009$, performance drops to 67%. This illustrate the difficulties encountered when classifying across years, and is present in all across-year classification tasks. One fundamental problem is that the data does not comply with the IID assumption of the learning algorithm due to differences in reef appearance. The third experiment, $2008 + 2009 \Rightarrow 2010$, shows an apparent gain in accuracy (83 %) but much of this is can be attributed to the increase in the random classification baseline. On the sub-classification tasks we note the excellent ability of this method to discriminate corals vs. non-corals: 95% (random assignment achieve up to 75%). Even stronger results are achieved on the coral discrimination task (given that we know it's a coral), with up to 97% accuracy. The non-corals discrimination task is more challenging and hence shows more confusion, in particular among the algae classes. Figure 5 shows full confusion matrices for each experiment. Note how the main source of confusion is between CCA and each of the other classes respectively. Part of this confusion is likely because CCA overgrows dead coral; while the coral structure retains it shape it now classifies as CCA.

**Coverage results:** Keeping the application domain in mind we also consider how well the coverage statistics can be estimated. For this purpose we generated scatter plots showing, for each category, the coverage based on the computer generated annotations compared to the coverage based on the expert annotations for the $2008 + 2009 \Rightarrow 2010$ experiment (Fig. 6). The first sub-plot shows coral coverage, where we consider the binary coral vs. non-coral scenario. The slope coefficient, $p1 = 1.01$, indicates a very robust ability to estimate coral coverage even for this challenging across-years experiment. This is important since many surveys, in particular those carried out over large spatio-temporal scales by reef managers and government agencies, often focus on this one metric (coral coverage) as an indicator of reef health. Coverage of individual coral genera *Porites, Pocillopora, Acropora* and *Pavona* are also accurately estimated while *Montipora* is underestimated. This might be related to the encrusting morphology of this genus which makes it less visually distinctive than the branching corals. The algal plots exhibit more variation.

| | **2008 $\Rightarrow$ 2008** | **2008 $\Rightarrow$ 2009** | **2008, 2009 $\Rightarrow$ 2010** |
|---|---|---|---|
| **Full** | 0.74 (0.28) | 0.67 (0.29) | 0.83 (0.39) |
| **Coral, non-corals** | 0.92 (0.66) | 0.93 (0.70) | 0.95 (0.75) |
| **Within corals** | 0.97 (0.35) | 0.91 (0.37) | 0.97 (0.43) |
| **Within non-corals** | 0.78 (0.43) | 0.70 (0.42) | 0.87 (0.52) |

Table 2. **Result details:** For each experiment we report the accuracy on the full 9-class experiment, as well as three key sub problems. The first 'coral vs. non-corals', considers the binary classification task where all corals are grouped together against non-corals. The second, 'within corals', considers the situation where we know a certain point is a coral, and we need to assign a genus level label. The third, 'within non-corals', consider the situation where we know a certain point is a non-coral and we need to assign the correct label (CCA, Turf algae, Macroalgae, Sand) . The numbers in parenthesis show the accuracy by random assignment of labels according to the priors.

**Discussion:** As discussed in Sec. 2, the method accuracy must be evaluated with respect to the inconsistent nature of the 'ground truth' annotations. Compared to the CRED report discussed in Sec. 2 the results in Table 2 seem very encouraging indeed, but since the CRED effort analyzed a different reef, used a different label set and used non-expert annotators, one must regard this comparison with caution. Again, we emphasize the importance of a large




(a) (b) (c)

Figure 5. **Confusion matrices:** The rows correspond to the true class assignments, the columns to the predicted assignments. A perfect performance would generate a matrix with ones on the diagonal, and zeros elsewhere. The numbers on the right side show how many samples are in the test set for each category. Note that some of the coral classes are rare when compared to the larger CCA class. The main source of confusion is between CCA and each of the other classes respectively, while there is little confusion between the coral classes

inter-operator variability study to better understand 'how good is good' for this task. Other future work will include a more careful evaluation of the challenges posed by color correction under water, as well as an appropriate treatment of the non-IID data situation.

## 5. Conclusion

We introduce coral reef image annotation as a relevant and important application domain to the computer vision community and propose a benchmark dataset. We argue that an effective treatment of this problem could have an impact on many disciplines such as landscape surveying and remote sensing and that our data set provides an excellent opportunity for the evaluation of computer vision methods. We also propose a novel method for object recognition in this setting that offers a strong baseline on the MLC dataset. This method handles ambiguous and 'organic' object boundaries by a multiple scale approach, which is superior to using any single patch size, and we show that careful treatment of color information further boosts performance. The proposed method yields $83.1\%$ accuracy on the MLC $2008 + 2009 \Rightarrow 2010$ experiment. We also show that the proposed method accurately estimates coral coverage across reef sites and multiple years, which offers exciting potential for large scale coral reef analysis.

**Acknowledgements:** The authors like to thanks Vincent Moriarty at California State University Northridge, for the annotation work that made this study possible, and Tali Treibitz at University of California, San Diego for valuable advice on the treatment of color information. This work was partly supported by NSF Computer Vision Coral Ecology grant #ATM-0941760. The image data was collected, curated, and analyzed with funding from grants OCE 04-17413 and OCE 10-26851 from the National Science Foundation and gifts from the Gordon and Betty Moore Foundation. This is a contribution of the Moorea Coral Reef (MCR) LTER, and is contribution number: 179 of the CSUN Marine Biology Program.

## References

[1] E. Adelson. On seeing stuff: the perception of materials by humans and machines. In *SPIE*, 2001. 2

[2] G. W. C. C. Kanan, A. Flores. Color constancy algorithms for object and face recognition. In *IEEE ISVC*, 2010. 4

[3] J. H. Carleton and T. J. Done. Quantitative video sampling of coral reef benthos: large-scale application. *Coral Reefs*, 1995. 4

[4] N. Carlevaris-Bianco, A. Mohan, and R. Eustice. Initial results in underwater single image dehazing. In *OCEANS*, 2010. 4

[5] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM TIST*, 2011. 5

[6] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 1995. 5

[7] O. G. Cula and K. J. Dana. 3d texture recognition using bidirectional feature histograms. *IJCV*, 59:2004, 2004. 2

[8] P. F. Culverhouse, R. Williams, B. Reguera, V. Herry, and S. Gonzales-Gil. Do experts make mistakes? a comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, 2003. 1, 3

[9] K. J. Dana, B. Van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real-world surfaces. *ACM TOG*, 18, 1999. 2, 4

[10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 2010. 2

[11] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *CVIU*,

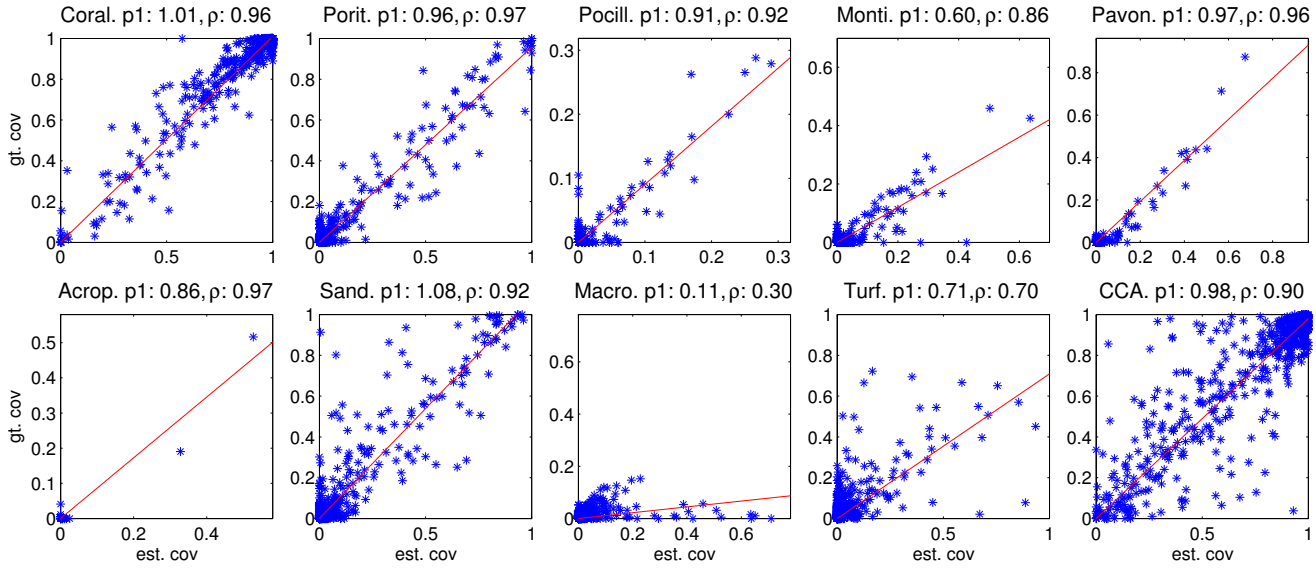



Figure 6. **Coverage estimation scatter plots:** for the $2008 + 2009 \Rightarrow 2010$ experiment. Each dot indicates the ground truth and the estimated coverages respectively per image. For each label type we performed linear interpolation (constrained to zero intersect) and report slope $p1$, and the Pearson product-moment correlation coefficient, $\rho$. A perfect estimation would generate numbers $p1 = 1, \rho = 1$. The first plot is an aggregated plot showing the coverage estimation for the binary coral vs. non-coral classification task.


2007. 2

[12] Personal correspondence with Matthew Dunlap, NOAA Coral Reef Ecosystem Division, Honolulu, HI, 2012. 4

[13] J. A. Hartigan and M. A. Wong. A K-means clustering algorithm. *Applied Statistics*, 1979. 5

[14] E. J. Hochberg and M. J. Atkinson. Capabilities of remote sensors to classify coral, algae and sand as pure and mixed spectra. *Remote Sensing of Environment*, 85, 2003. 1

[15] J. C. Kenyon, R. E. Brainard, R. K. Hoeke, F. A. Parrish, and C. B. Wilkinson. Towed-diver surveys, a method for mesoscale spatial assessment of benthic reef habitat: A case study at midway atoll in the hawaiian archipelago. *Coastal Management*, 2006. 1

[16] K. E. Kohler and S. M. Gill. Coral point count with excel extensions (CPCe): A visual basic program for the determination of coral and substrate coverage using random point count methodology. *Computers & Geosciences*, 2006. 2

[17] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *IJCV*, 2001. 4

[18] N. MacLeod, M. Benfield, and P. Culverhouse. Time to automate identification. *Nature*, 2010. 1

[19] J. Malik, S. Belongie, J. Shi, and T. Leung. Textons, contours and regions: Cue integration in image segmentation. In *ICCV*, 1999. 2

[20] J. Malik, S. Belongie, J. Shi, and T. Leung. Textons, contours and regions: Cue integration in image segmentation. In *ICCV*, 1999. 4

[21] M. S. A. Marcos, M. Soriano, and C. Saloma. Classification of coral reef images from underwater video using neural networks. *Opt. Express*, 2005. 2

[22] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *PAMI*, 2004. 4

[23] R. Ninio, S. Delean, K. Osborne, and H. Sweatman. Estimating cover of benthic organisms from underwater video images: variability associated with multiple observers. *Marine Ecology Progress Series*, 2003. 3

[24] M. Patterson and N. Relles. Autonomous underwater vehicles resurvey bonair: a new tool for coral reef management. In *ICRS*, 2008. 1

[25] T. Randen and J. Husoy. Filtering for texture classification: A comparative study. *PAMI*, 2002. 2

[26] F. Rohwer and M. Youle. *Coral Reefs in the Microbial Seas*. Plaid Press, 2010. 1

[27] M. Shiela, A. Marcos, L. David, E. Peñaflor, V. Ticzon, and M. Soriano. Automated benthic counting of living and nonliving components in ngedarrak reef, palau via subsurface underwater video. *Environmental Monitoring and Assessment*, 2008. 2

[28] M. Stokes and G. Deane. Automated processing of coral reef benthic images. *Lim. Ocea. Methods*, 2009. 2

[29] A. Torralba and A. A. Efros. Unbiased look at dataset bias. In *CVPR*, 2011. 2

[30] T. Treibitz and Y. Schechner. Active polarization descattering. *PAMI*, 2009. 4

[31] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *IJCV*, 2005. 2, 4

[32] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *PAMI*, 2009. 2

[33] J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *ICCV*, 2005. 2