Classification of Marine Organisms in Underwater Images using CQ-HMAX Biologically Inspired Color Approach

Sepehr Jalali, Paul J. Seekings, Cheston Tan, Hazel Z. W. Tan, Joo-Hwee Lim and Elizabeth A. Taylor

Abstract—In many coastal environments, particularly in tropical zones, coral reef ecosystems have exceptional biodiversity, contribute to coastal defense, provide unique and important habitats and valuable commercial resources. Assessment of environmental impacts on biodiversity in such areas are increasingly important to mitigate potential adverse effects on specific ecosystems. Visual classification of marine organisms is necessary for population estimates of individual species of corals or other benthic organisms. In this paper, we introduce a new image dataset of benthic organisms that are of different colors, shapes, scales, visibility and are taken from different viewpoints. We evaluate several different classification approaches on this dataset, and show that CQ-HMAX, our new biologically inspired approach to utilizing color information for object and scene recognition, that is inspired by the characteristics of color- and object-selective neurons in the high-level inferotemporal (IT) cortex of the primate visual system, results in better classification results in comparison with existing computational models such as support vectors machines, SIFT based approaches and the HMAX biologically inspired approach. We show that concatenating our model which encodes color information with the HMAX model which encodes grayscale shape information results in the highest classification accuracy.

I. INTRODUCTION

Coral reef ecosystems are critically important for the health of some coastal environments and this particularly applies in the tropics where coral reefs contribute to coastal defence against physical impacts such as storms, provide habitats suitable for juvenile fish and other marine organisms that require sheltered environments, and create valuable ecotourism and other commercial opportunities. Environmental impact studies are increasingly required before coastal development is approved, and in South-east Asia where approximately 70 percent of the population lives in coastal zones, this is a significant factor with trans-boundary issues adding to its importance. However, assessment of the health of coastal habitats - which in essence means studying the variety of species in a given area, their abundance and their health - is time-consuming and expensive, and very often not possible on any but locally focused scales. It follows that a way of facilitating this process would be welcomed by the scientific community and would also be cost-effective in application.

Coral reefs and other benthic marine organisms in the tropics contribute to the wide diversity of species in the benthic ecosystem [1]. They contribute to improved water quality and visibility in littoral waters. Suspension feeders form the base of primary food production and are an integral link between higher and lower levels of complex food webs [2]. Due to this, many benthic marine organisms play a major role in carbon cycling as mentioned in [3].

Population level estimates of the relative abundance of individual species of coral or other benthic organisms are important in assessing overall coral health and environmental impacts. Such surveys are either executed in situ by by teams of trained divers or via collection of reef images acquired in the form of stills or via video transects. In the case of the former, “permanent quadrats” are marked out on the reef using markers at four corners of the chosen area and photographed at each sampling time. This method supports low variation arising from inconsistent placement of the sampling unit. Video transects lend themselves to large-scale surveys of overall reef condition, and in many areas have replaced methods such as manta-tows which often result in high observer error. In both cases, post-processing of the photos, i.e. identification and quantification of the area, is the greatest rate-limiting factor and is extremely time consuming. A good classification algorithm could significantly speed up this process. In this paper we use images which do not require segmentation to demonstrate the ability of HMAX and CQ-MAX for the classification stage.

Several visual recognition models are inspired by the hierarchical organization of the visual cortex, such as [4] and [5]. Most of these models focus on grayscale information and ignore color information. While the broad use of color information in the primate visual system is well-known, the details are still under active investigation [6]. HMAX is a biologically-inspired model which focuses on the shape processing capabilities of the ventral visual pathway, and has been used to perform classification tasks [7]. Our model focuses on modeling the high-level use of color by incorporating insights from cognitive psychology and neuroscience. The broad intuitive inspiration for our model follows from the fact that colors are recognized categorically just as object classes are, even though color discrimination and matching is continuous [8]. Interestingly, people of different races [9], as well as chimpanzees [10], organize colors into the same basic color categories, such as red, blue, yellow, green.

Overall, our model can be loosely described as performing object and scene classification by reducing a given image to a “coarse arrangement of categorical color blobs”, similar to the idea of spatial aggregation of visual keywords [11], but with realization on the HMAX model. This is different from
approaches that utilize color information in a low-level fashion, although the two approaches are not mutually exclusive. Crucially, our biologically-inspired approach outperforms the naive use of color where an image is decomposed into separate color channels that are processed independently until the final classifier stage as shown in [12].

In this paper, we present a simplified overview of information processing in the human and dolphin visual cortex, with emphasis on the role of color processing in Section II. We present a brief review of the HMAX model [5] and review our biologically inspired color model presented in [12] in Section III. We introduce a new dataset of images of marine benthic organisms and show the experimental results of applying different methods on this dataset in Section IV followed by a discussion in Section V and conclusions and future work in Section VI.

II. Color Vision in Humans and Marine Mammals

The term color vision refers to the capability of a visual system to respond systematically to light differing in wavelength only. There have been relatively few studies of aquatic mammals. Bottlenose dolphins have been reported to be effectively color blind, seeing only variations in brightness [13]. However, more recent work has shown that a bottlenose dolphin could discriminate between 397 nm and 487 nm (violet and blue light), but not between 457 nm and 544 nm (violet-blue and green) which suggests that they do possess a mechanism for color discrimination [14]. This wavelength bias presumably corresponds to an optimal operating range driven by evolutionary requirements. Similarly, behavioral studies have shown that seals and California sea lions exhibit maximum sensitivity in the green part of the spectrum which might indicate adaptation to a specific underwater environment. While little is known about the color vision in species that have evolved to live in tropical coastal waters, it seems likely that multiple pigment visual systems in marine mammals probably serve three functions:

1. to maximize sensitivity at depth and approaching prey from below; 2. to maximize contrast when at shallower depths and approaching prey from the side; and, 3. to allow vision in up to nine orders of magnitude of changes in light intensity (bright light near the surface would bleach sensitive, dark-adapted retinal rods).

HMAX simulates the simple-complex cell hierarchy in the visual cortex of humans and primates. The model reflects the general organization of the visual cortex as a series of distinct brain areas from primary visual cortex (a.k.a V1) to inferotemporal cortex (ITC). There are four layers in the model (namely: S1, C1, S2 and C2). The S layers serve to increase the selectivity of the system by performing template-matching, while the C layers serve to increase the position and scale invariance of the system by performing pooling over positions and scales.

The general increase in selectivity and invariance up the hierarchy of the visual system has been well-documented [15]. More specifically, evidence for the MAX-like (or OR-like) pooling used by the model has been found in cats and monkeys [16], [17], [18]. Meanwhile, evidence for conjunctive (AND-like) template matching used by the model has also been found in macaque monkeys using single-neuron recording techniques. Specifically, it was found that neurons in area V2 are sensitive to combinations of orientations [19]. Since area V2 receives inputs from V1 (which contains neurons analogous to units in layers S1 and C1), these cells are analogous to units in model layer S2. Neurons in ITC, which is the highest layer in the brain’s visual pathway that is purely visual, have been found to contain object information that can be decoded in a manner that is invariant to position and scale. Such cells are thus analogous to units in model layer C2.

The previous studies quoted above have focused on investigating shape processing, and therefore did not investigate color processing. However, it is clear that the brain’s visual pathway is also sensitive to color (see [20] for a review). Since it is very well-established that lower levels of the visual pathway are involved in color processing [21], here we focus specifically on the high-level ITC.

A majority (70%) of ITC neurons have been found to be color-selective [22], [23], [24]. More direct and causal evidence for the role of ITC in color processing has been found using brain-lesion experiments [25], brain cooling [26] and position emission tomography (PET) imaging [27]. Color-selective neurons in ITC are found in clusters, suggesting that they may form a roughly segregated and independent processing network [28], [29], [30]. Different ITC cells prefer different colors, and the preferred colors collectively span most of the color space [23], [29]. Nonetheless, ITC neurons tend to be selective for the basic colors [20], [31], [32], i.e. “red”, “blue”, “yellow” and “green”. Finally, the region of ITC where color-selective neurons are found is coarsely retinotopic [33], meaning that there is only coarse spatial information.

We developed and implemented a new approach to color encoding, inspired by the characteristics of color and object selective neurons in the high level inferotemporal (IT) cortex of the primate visual system in [12]. In our hierarchical model, we introduce a dictionary of features that encode quantized color patches, along with coarse, relative spatial information. We implement this model on a new dataset of images of different marine benthic organisms taken underwater from different ranges, viewpoints and visibility and find that the performance of this color model in concatenation with various shape features results in best classification performance on our dataset. While we implemented this approach by extending one specific model (HMAX), the approach can also be incorporated into other models.

III. CQ-HMAX Model

In this section, we present a brief overview of our biologically inspired model, CQ-HMAX (Color Quantization
Hierarchical Max) which uses color information in a hierarchical organization of simple and complex cells which is described in more detail in [12], [34]. The combination of HMAX model with CQ-HMAX is investigated and the final model which encodes both color and shape information is presented.

HMAX is a hierarchical model which uses Gabor filters in the first layer and has interval simple and complex cells to extract and pool features in the images. CQ-HMAX model has a similar hierarchical structure. However, color quantization cores are used in it instead of Gabor filters, hence CQ-HMAX model encodes color information as shown in Fig. 1. When combined with HMAX, the overall model includes both color and shape information.

In CQ-HMAX, an image pyramid is created in YIQ color space. The pyramid has 10 scales, with each neighboring scale different by a ratio of $1/(2^{1/4})$. A set of representative values from each color channel is selected as color cores and used to find the best matching unit to each individual pixel value in the pyramid. The $S1$ layer is created on 10 scales indicating the index of the best matching YIQ core to each pixel in the image pyramids. The outputs at the $S1$ layer are the index values (i.e. 1,2,...,125) of the best-matching color core for each element in the image pyramid. At the $C1$ layer, a local max pooling is computed over $\pm 10\%$ spatial neighborhoods of approximately $6 \times 6$ on $\pm 1$ neighbor scales to find the most frequent color core in each neighborhood. The $C1$ layer provides local invariance to position and scale as it pools nearby $S1$ units, and as a result, subsamples $S1$ to reduce the number of units. Once the $C1$ layer is created, sampling is performed by centering patches of size $4 \times 4$ at random positions and scales using a normalized random number generator function to create a dictionary of features. A distance table is created to store the actual weighted Euclidean distances of the indices from YIQ quantization cores. Once the dictionary of features and the distance table are created, each entry in the dictionary of features is used as a filter to be convolved on $C1$ patches of size $4 \times 4$ on the neighbor scales of the dictionary feature in the pyramid. The responses $V(d,p)$ of each dictionary feature, $d$ to all of the neighbor patches of the same size in $\pm 1$ scale and $\pm 10\%$ in position, $p$ are calculated using a
Euclidean distance equation. Once the $S_2$ layer is generated, the maximum values for each patch in the dictionary are taken as the $C_2$ output. This layer outputs a vector of the same size as the dictionary of features. The $C_2$ vector is then fed to a SVM for classification.

A. Encoding Shape Information

In order to implement the use of shape information, we use the HMAX hierarchical model presented in [35] with the code provided in [36]. In the first layer, the normalized dot product of Gabor filters of different orientations are calculated over different sizes of the image pyramid ($S_1$ layer). In the $C_1$ layer a local max on neighbor positions and scales is taken for pooling in order to provide invariance to scale and position of features. A dictionary of features is sampled from $C_1$ layer outputs in two modes of random sampling and using frequency and spatial information of features (which will be further explained). Once the dictionary of features is created, the response of each feature in dictionary to each image patch is calculated ($S_2$) and a max is taken over all image patches and over all dictionary features in different combinations ($C_2$). An $N$-dimensional vector is calculated as the output of the $C_2$ layer, where each element is the maximum response (everywhere in the image in [37] and in a spatial neighborhood of each dictionary feature in [35]) over image patches for each dictionary feature where $N$ is the number of features in the dictionary. $C_2$ vector is then fed to a linear support vector machine for classification.

Our final model is based on the concatenation of CQ-HMAX and HMAX models as shown in Fig. 2.

B. SIFT Features

Scale Invariant Feature Transform (SIFT) feature extraction method [38] is a well-known method which has produced promising results in classification tasks. Here we investigate its application in classification of marine organisms. In SIFT methods, a series of features are calculated using difference of Gaussian (DoG) methods over different scales. Once a set of features are selected, features from new images are compared with these candidate regions using their Euclidean distance and from the full set of matches. A subset of key point features which agree on the object, its scale,
orientation and location in the new image, are identified to filter out good matches. Finally a histogram of features is calculated and the final histograms are sent to a SVM classifier. In this experiment we use PHOW features (dense multi-scale SIFT descriptors), Elkan k-means for fast visual word dictionary construction, spatial histograms as image descriptors, a homogenous kernel map to transform a $\chi^2$ support vector machine (SVM) into a linear one and finally an internal SVM for classification using VLFeat toolbox [39].

IV. MARINE ORGANISMS DATASET AND EXPERIMENTAL RESULTS

The marine benthic organisms dataset includes 19 classes of marine organisms that grow on or are closely related to the benthos (the seabed). Each class contains between 60 to 300 color images of different sizes. The size of each image is approximately 5000 × 3000 pixels. However, due to high computational costs, images are resized to approximately 500 × 300 pixels, while maintaining the aspect ratio. We used 30 randomly selected images for training from each class and the rest of the images were used in the test phase. Sample images of every class of this dataset are shown in Fig. 3.

Many benthic marine organisms have several distinguishing factors which set them apart from each other. Some visual characteristics of some of the classes used are as follows: boulder or submassive corals are easily differentiable from others by their roughly spherical shape which is similar in all dimensions except the base which is flattened. The foliate organisms have a leaf-like appearance with folded plates or spires extending upwards. Branching corals have an outward growth of branches which have primary and secondary branchings, unlike digitate forms which do not have secondary branches. The plate-like corals have laminar and flattened sheets which may be vertical or horizontal. Mushroom soft corals have a unique appearance of a flat uneven circle or oval. Anemones have a single body with tentacles radiating in all directions. With all these distinguishing factors and more, the model recognizes and is able to differentiate them, hence providing us with useful identifications [40]. In this dataset, we have 19 categories (and sub-categories) of benthic marine organisms: 1- Algae, 2- Anemone (Lily), 3- Anemone (Reef), 4- Body Sponge, 5- Boulder, 6- Branching, 7- Branching (Soft), 8- Digitate, 9- Encrusting, 10- Foliate, 11- Mushroom Coral, 12- Mushroom (Soft Coral), 13- Plate, 14- Seafan (Soft Coral), 15- Seagrass (Sargassum), 16- Zoanthids, 17- Seagrass (Seaweed), 18- Stem Sponges, 19- Tubulate.

We evaluated different classification methods on this dataset such as naive use of SVM in which images are resized to the size 160 × 90 pixels and sent to a linear kernel classifier. In other experiments, we used HMAX and SIFT methods described in Section III. As can be seen in Table I, the use of CQ-HMAX model provides significant improvements over using shape based HMAX model and when this model is concatenated with HMAX model, the highest accuracy is achieved. Feeding images directly to a support vector machine results in a very low classification accuracy of 20.5%. When HMAX model is applied on this dataset, and scale invariant features are extracted, a boost in classification is achieved to enhance it to 40%. A SIFT based method results in about 52% accuracy on this dataset. The use of CQ-HMAX results in better classification accuracy than SIFT and HMAX and reaches 56.2%. Concatenation of $\chi^2$ vectors of HMAX model (shape features) with CQ-HMAX model (color features) results in the highest classification accuracy of 61.2% over this dataset. These results are inline with the previous experiments carried out on the CQ-HMAX model in [12] where concatenation of HMAX and CQ-HMAX models results in a better classification accuracy in several datasets such as Flowers, Soccer, Caltech101 and Scenes. In this paper, concatenation of CQ-HMAX with SODO-HMAX model of Zhang et al. [41] results in classification accuracy better than the state of the art in Soccer [42] and Flowers [43] datasets.

TABLE I: Classification accuracy on the marine benthic organisms dataset using different methods.

<table>
<thead>
<tr>
<th>Classification Model</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>20.54%</td>
</tr>
<tr>
<td>HMAX (Grayscale)</td>
<td>39.61%</td>
</tr>
<tr>
<td>HMAX (Red Channel)</td>
<td>40.94%</td>
</tr>
<tr>
<td>HMAX (Green Channel)</td>
<td>39.92%</td>
</tr>
<tr>
<td>HMAX (Blue Channel)</td>
<td>41.13%</td>
</tr>
<tr>
<td>HMAX (RGB Channels)</td>
<td>48.22%</td>
</tr>
<tr>
<td>SIFT</td>
<td>52.11%</td>
</tr>
<tr>
<td>CQ-HMAX</td>
<td>56.23%</td>
</tr>
<tr>
<td>HMAX + CQ-HMAX</td>
<td>61.18%</td>
</tr>
</tbody>
</table>

V. DISCUSSION

By the use of color information alone, we could achieve a higher classification accuracy than the use of shape information alone. The use of color and shape information in this classification task results in the best performances achieved. As can be seen in Fig. 4, CQ-HMAX (red bars) outperforms HMAX model (blue bars) in almost all classes.
One of the classes in which CQ-HMAX significantly improves over HMAX are class 17 and 18 (Seagrass Seaweed and Stem Sponges) where images have similar colors, but are of different viewpoints, scales and orientations. A few samples of this class are shown in Fig. 5a,c. On the other hand, in classification of images of Seafan category (Class 14) shown in Fig. 5b, HMAX performs slightly better than CQ-HMAX and this is due to the variety of colors of images in Seafan category and the consistency of oriented edges. In most of the other classes, CQ-HMAX performs as well or better than HMAX. Class 2 (Lily Anemone) has the highest classification accuracy in HMAX model, and the accuracy is equally as good as CQ-HMAX. This is due to consistency in the oriented bars, and existence of enough training samples from different colors of this class.

As Table I shows, the use of SVM on the raw pixels does not result in a high classification accuracy and this is due to the different intraclass viewpoints, scales and varieties in the colors. Hence a model that is more invariant to the viewpoints and scales would match this dataset better. SIFT based methods and HMAX model, both provide invariance to scale and position of the features, however they are also selective to the intraclass variations when the orientations of the edges are very random among images in the same class.

Since this is a dataset of live organisms which generally do not have a firm rigid structure, and the images are taken from different viewpoints, these models do not provide the highest classification accuracy. On the other hand, CQ-HMAX model is very invariant to changes in the rotations of the edges in the images and encodes the color information which is an important characteristic in this model. As Table I shows, the use of R, G and B channels separately does not result in any significant improvement but their combination results in a better classification accuracy which is below the CQ-HMAX model as HMAX model is more orientation based. As Fig. 4 shows, the classification accuracy achieved with HMAX
model is highest in Class 2 (Fig. 5d) where the orientation of lines in the images are consistent.

Another interesting characteristic of these two models is that despite having a similar accuracy in many classes, concatenation of the two models results in a better classification.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we introduced a new dataset of marine benthic organisms and performed different classification methods to distinguish images from one another. We showed that using color information (CQ-HMAX model) along with shape information (HMAX) results in the best classification results and outperforms the SIFT based classification and the naive use of SVM.

As can be seen in Fig. 1, the $S_1$ and $C_1$ layers resemble a segmentation of the images, which shows the potential use of CQ-HMAX model for a segmentation task on this dataset.

One of the future applications of this model is to enable an automatic system for classification of marine organisms underwater in certain areas to investigate their abundance. Since the classification accuracy achieved with our model is very reasonable (about 61%) and in real life situations the number of classes in a specific area are often fewer than the classes covered in this dataset, this accuracy will probably improve and hence this model could be used for segmenting the images and classifying the marine organisms into broad classes.

One of the basic concepts in CQ-HMAX model is the color quantization. In this model, a series of cores are hard-coded in the YIQ color space. However, the choice of quantization cores can be modified in future extensions of this model to better match the vision system of the aquatic mammals especially when dealing with low quality underwater images which are very common due to the particulate matter in the water column. While dolphins and sea lions are able to see in air, vision in these mammals has adapted to life underwater where discrimination within the blue/green/grey spectrum is most useful and where contrast discrimination is important. This can be further explored to guide the choice of different quantization cores.

Dead, bleached corals and healthy live ones are a crucial deciding factor in assessing the health of coral reef ecosystems so differentiating between them is important. In future improvements of this model, this would be an additional factor.

Color changes due to fluorescence in some marine organisms under different lighting conditions is another area we
are studying that might have several applications including assessment of the health of reef associated organisms.

Acknowledgements

The photos were taken by Jim Wong who was working for the Marine Mammal Research Laboratory, Tropical Marine Science Institute, NUS.

REFERENCES