A Deep Learning Paradigm for Detection of Harmful Algal Blooms

Arun CS Kumar Suchendra M. Bhandarkar
Department of Computer Science, The University of Georgia, Athens, GA 30602-7404, USA
E-mails: {aruncs@uga.edu, suchi@cs.uga.edu}

Abstract

Effective and cost-efficient monitoring is indispensable for ensuring environmental sustainability. Cyanobacterial Harmful Algal Blooms (CyanoHABs) are a major water quality and public health issue in inland water bodies. The recent popularity of online social media (OSM) platforms coupled with advances in cloud computing and data analytics has given rise to citizen science-based approaches to environmental monitoring. These approaches involve the lay community in the acquisition, collection and transmission of relevant data in the form of tweets, images, voice recordings and videos typically acquired using low-cost mobile devices such as smartphones or tablet computers. While cost effective, citizen science-based approaches are highly susceptible to noise, inaccuracies and missing data. In this paper we address the problem of automated detection of harmful algal blooms (HABs) via analysis of image data of inland water bodies. These image data are acquired using a variety of smartphones and communicated via popular OSM platforms such as Facebook, Twitter and Instagram. To account for the wide variations in imaging parameters and ambient environmental parameters we propose a deep learning approach to image feature extraction and classification for the purpose of HAB detection. The current system is a first step in the design of an automated early detection, warning and rapid response system that can be adopted to mitigate the detrimental effects of CyanoHAB contamination of inland water bodies.

***Keywords*—** deep learning, texture classification, image segmentation, HAB detection, citizen science.

1. Introduction

Accurate, cost-effective, and targeted monitoring is a critical aspect of sustainable management of the environment. *Cyanobacterial Harmful Algal Blooms* (CyanoHABs) are a major water quality and public health issue in inland water bodies, such as lakes and ponds, since they hamper recreational activities, degrade aquatic habitats and potentially affect human, livestock and wildlife health via toxic contamination [2, 6, 9, 37]. Existing approaches for monitoring of inland water bodies, i.e., *in situ* sampling and satellite-based remote sensing, suffer from serious limitations in terms of cost, data resolution and accuracy. Traditional *in situ* sampling techniques are laborious and time-consuming and prohibitively expensive for real time monitoring of high-frequency environmental phenomena such as water pollution and natural hazard impact assessment. Satellite-based remote sensing techniques can be cost effective but are much less accurate because of mixed pixel issues, geometric noise and radiometric noise. The inherent limitations on the resolution of satellite image data often introduce errors in the resulting prediction models and algorithms. *In situ* remote sensing, using hyperspectral sensors deployed in the field, addresses some of the limitations of satellite image data. Hyperspectral sensors can be useful in rapid, targeted, and cost-effective monitoring of environmental phenomena, especially when these sensors operate, acquire, and transmit data based on an observer-induced triggering mechanism. However, hyperspectral sensors typically have a limited spatial field of view and hence preclude a dense spatial sampling of the underlying environmental phenomena. This may cause the sensors to miss certain environmental phenomena, such as CyanoHABs, in their early stages of development.

With the increasing popularity of online social media (OSM) platforms, such as Facebook, Twitter and Instagram, coupled with advances in cloud computing and data analytics, there has been a growing trend towards adopting citizen science-based approaches to environmental monitoring. The goal is to leverage the OSM platforms to expand the observer base to include not only environmental scientists and restoration officials but also the lay community at large (i.e., citizen scientists), which includes local area residents and businesses, fishermen and tourists, to encourage more frequent and comprehensive environmental monitoring. The citizen science-based approaches are typically based on acquisition, collection and transmission of data, in the form of tweets, images, voice recordings and videos, using low-cost mobile devices such as smartphones or tablet computers. The citizen science-based approaches are intended to com-

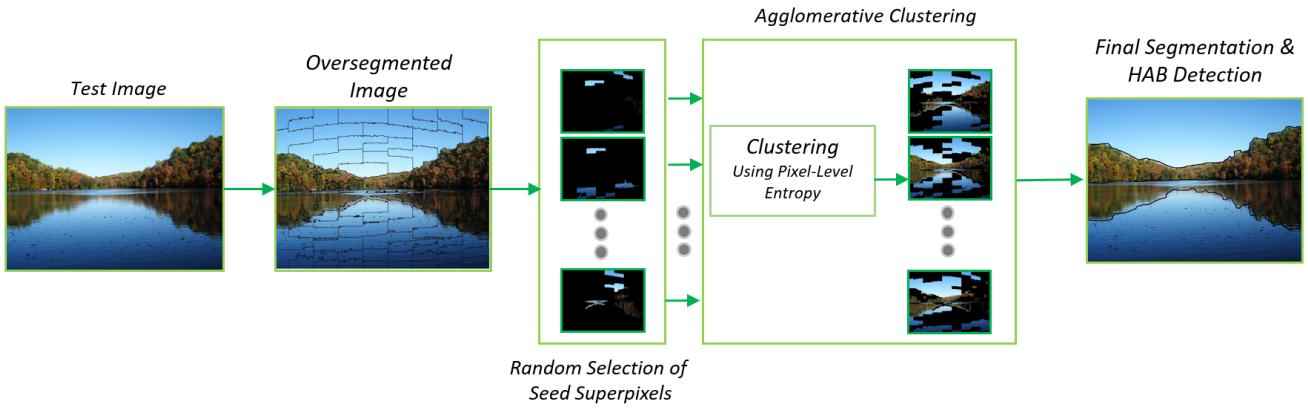

Figure 1. Pipeline of the proposed framework for HAB detection

plement the traditional *in situ* and remote sensing-based approaches to environmental monitoring.

In this paper, we describe the design and implementation of a citizen science-based platform for early detection of CyanoHABs in inland water bodies. In particular, we address the problem of automated detection of harmful algal blooms (HABs) via analysis of images of inland water bodies. The image data are acquired using a variety of smartphones and tablet computers and communicated via popular OSM platforms such as Facebook, Twitter and Instagram to a central data server. One of the primary challenges in analyzing these images is ensuring robustness of the resulting algorithms to the wide variations in imaging parameters (i.e., camera resolution, viewpoint, scale and clutter) and ambient environmental conditions (i.e., weather and illumination). To this end we propose a deep learning-based approach to image feature extraction and classification for the purpose of HAB detection. We demonstrate the effectiveness of the proposed approach via experimental results on a diverse image dataset comprising of images of various inland water bodies acquired using a variety of smartphone cameras with wide variations in imaging parameters and ambient environmental conditions. With the robust integration of multimodal data (i.e., tweets, images, videos, voice, and text) from localized detrimental environmental phenomena, we expect the proposed system to constitute a first step towards the development of an early warning, rapid response and optimal sensor deployment infrastructure for comprehensive environmental monitoring, sustainability and restoration.

2. Related work

Most recent work on the detection of HABs in inland water bodies has focused on analysis of remotely sensed satellite images where the goal is to detect and map CyanoHABs based on the phycocyanin (PC) absorption characteristics of the reflected spectrum in the 620 nm wavelength band [18]. The 620 nm reflectance band is typically contaminated by chlorophyll-*a* (found in all phytoplankton and green algae), colored dissolved organic matter (CDOM) and totally suspended sediment (TSS) in the water body [3]. Consequently, extracting the characteristic CyanoHAB spectral signature entails a complex deconvolution operation [19]. To this end, several algorithms have been proposed for quantification of cyanobacterial concentration from satellite data: the single band ratio algorithm [32], semi-empirical algorithm [7], nested semi-empirical band ratio algorithm [35] and the maximum peak height (MPH) algorithm [20]-[23]. These algorithms are also used to compute various water quality indices. Matthews and Odermatt [23] use the MPH algorithm to compute the sun-induced chlorophyll fluorescence (SICF) peak, the sun-induced phycocyanin absorption and fluorescence (SIPAF) peak, the normalized difference vegetation index (NDVI) and the backscatter and absorption-induced reflectance (BAIR) peak. Oyama et al. [27] compute the maximum chlorophyll index (MCI), cyanobacterial index (CI), floating algal index (FAI), normalized difference vegetation index (NDVI) and the normalized difference water index (NDWI) using a variant of the MPH algorithm.

Although satellite remote sensing can be used to monitor environmental parameters across a broad study area at a regional scale, it has serious practical limitations when monitoring a localized phenomenon such as the initiation of CyanoHABs in a water body. Satellite data cannot be used to develop an accurate early warning system for localized environmental phenomena because of their limited spatial, spectral, and temporal resolution. Issues such as mixed pixels, atmospheric interference, lack of a useful band center, and occasional poor site revisit time could potentially introduce serious errors in the predictive model or cause it to miss the environmental phenomenon altogether. *In situ* remote sensing using hyperspectral sensors mounted in the vicinity of the water body can address some of the short-

comings of satellite remote sensing [24]-[26]. However, the limited field of view of these sensors precludes dense spatial sampling of large water bodies.

Citizen science-based monitoring has the potential to effectively complement the aforementioned traditional approaches to environmental monitoring. Image data acquired by lay citizens using a variety of smartphones and tablets and communicated via popular OSM platforms such as Facebook, Twitter and Instagram could be potentially used as input to an early CyanoHAB warning system. A major shortcoming of using such data is that they can be extremely noisy and also exhibit wide variations with respect to the imaging parameters (i.e., camera resolution, viewpoint, scale and clutter) and ambient environmental conditions (i.e., weather and illumination). In this paper, our goal is to perform early detection of HABs in inland water bodies using such image data acquired by lay citizens.

To ensure a low false alarm rate and low miss rate for the detection of HABs, the extracted image features and the image analysis algorithms need to be robust to noise and the aforementioned variations in imaging parameters and ambient environmental conditions. In recent times, deep (i.e., multi-layer) convolutional neural networks (CNNs or *ConvNets*) have been observed to be very effective in extracting robust high-level features (i.e., abstractions) from very diverse input image data [15]. In particular, generic image features generated by a deep *ConvNet* pretrained on the very diverse ImageNet dataset [31] have been shown to be capable of tackling a wide range of image recognition tasks such as object detection, fine-grained object classification, scene recognition, attribute detection and content-based image retrieval on very diverse datasets [34].

In the context of HAB detection, image data acquired by lay citizens are observed to contain a lot of clutter. Images may contain regions or surfaces other than those that can be classified as *lakes* or *ponds*, such as, *sand, sky, trees,* and *grass*, to cite a few. Since region classification is highly sensitive to noise and clutter, these noisy regions can seriously impact the HAB detection accuracy. To address this shortcoming, we propose a two-stage classification pipeline wherein the first stage entails a general classification procedure to extract *lake* regions from the input image using a bottom-up agglomerative clustering procedure. This is followed by the second stage which comprises of a fine classification procedure to detect HABs in the extracted *lake* regions. We rely primarily on *texture* and *color* cues for region classification by leveraging the deep *ConvNet* features. Textural properties provide useful information for understanding scenes and discriminating between foreground objects and the background, especially when the shapes of the objects do not convey useful information.

The computer vision research community has invested significant research effort in texture recognition and classification [5, 17, 33]. Most of the works on texture recognition and classification have traditionally focused on material recognition [33, 17] and scene understanding [29]. In contrast, recognizing regions or surfaces in cluttered natural scenes containing multiple object categories (such as *cloud, sky, water* etc.) is relatively less explored. Color features, on the other hand, have been observed to perform well for object detection [10], image classification [11] and texture classification [14]. Khan et. al. [12, 14] have shown that color attributes combined with textural cues result in a significant improvement in the performance of texture-based region classification and recognition algorithms.

3. Contribution of the paper

Our primary goal is to detect and segment HABs in an input image. To accomplish this goal, we first perform a general classification of surfaces or regions in the input image to extract the *lake* regions, followed by a domain-specific classification of the *lake* regions. This problem can be formulated as a general surface/region categorization problem, where the main challenge is that, surfaces in natural scenes are typically heterogeneous in terms of their color and texture attributes and highly cluttered. Recognizing surfaces or regions in natural scenes is particularly challenging on account of three reasons:
(1) It is extremely difficult to represent surfaces from natural images in appearance space on account of high intra-class variation. Textures of surfaces in natural images tend to appear as clutter in that they cannot be represented using a single texture category. For example, the surface of a *mountain* could be characterized as either *rocky*, *bushy*, *rugged* or more than one or all of the above (i.e., cluttered) and hence surface extraction using a single classifier would be highly inefficient.
(2) In some cases, lack of identifiable patterns renders texture recognition approaches impractical. For example, surfaces such as *sky* or *water* may not have any reliable textural attributes, which makes recognition challenging.
(3) Lack of high level cues such as shape, that could potentially aid recognition. Exploiting the global shape of an object has been shown to improve texture recognition performance in general [5]. However, in natural scenes many objects do not have a well-defined rigid shape.

The main contribution of this paper is the formulation of an *agglomerative clustering*-based optimization framework, inspired by [36], to integrate surface fragments or superpixels [30] such that the likelihood of surface categories that are too heterogeneous to be characterized by a single textural attribute is maximized. We define a heterogeneous surface category as a combination of multiple textural attributes, such that any subregion or superpixel of the entire heterogeneous surface, provides only a local view of the underlying surface category. It is only by combining or inte-

grating multiple such subregions or superpixels, in a manner that maximizes the likelihood of the surface category, that we can recover the entire underlying surface reliably. Commonly used segmentation techniques such as Conditional Random Field (CRF) models use edge potentials to exploit similarities between regions or pixels. However in our case, the goal is to group superpixels such that the likelihood of the object increases when grouped than not. We chose agglomerative grouping-based techniques as they are simple, parameter free and, being insensitive to the choice of number of initial segments, are shown to perform well in such situations [36]. The proposed formulation can be viewed as similar to putting jig-saw puzzle pieces together, since the puzzle pieces once properly joined make more sense than before. For the purpose of texture recognition we extend the *deep filter banks* technique proposed by [4]. We improvise the approach proposed in [4] by omitting the fully-connected (FC)-CNN descriptor which provides object shape information (since we do not rely on the high-level shape description of the underlying surface) while retaining the Fisher vector (FV)-CNN descriptor.

In the proposed framework, we take a top-down approach by over-segmenting the image to generate superpixels and cluster the superpixels iteratively based on texture, color and semantic contextual cues. The pipeline of the proposed approach is shown in Figure 1. We integrate textural and color attributes along with other learned contextual cues, to address the challenges mentioned above. The combination of textural attributes with color-based attributes has been shown to yield significant improvement in texture recognition performance [12, 14]. Semantic contextual cues are represented by pairwise spatial locations of surfaces learned during training, and have been shown to improve object recognition performance [8]. Since natural scenes are highly cluttered and contain textureless surfaces such as *sky* or *stagnant water*, incorporating semantic context into the proposed framework allows us to prune a lot of noisy classifications while recovering some of the textureless surfaces.

4. Overall System Description

The primary goal of the system is to perform early detection of HABs in inland water bodies using images acquired and communicated by lay citizens via popular OSM platforms such as Facebook, Twitter and Instagram. We have recently developed a mobile software application (i.e., app) that enables users to capture using smartphones and upload over the Internet, images of lakes across the world using the aforementioned OSM platforms. The mobile app also lets users provide a textual description of the lake, along with its *Global Positioning System* (GPS) coordinates and other relevant metadata, which is then published as a *tweet* upon submission. The purpose of the mobile app is to collect data as well as to create general public awareness about the seriousness of CyanoHAB contamination via publication of tweets on OSM platforms (in our case, on Twitter using a *hashtag*).

Data collection on a global scale from lay citizens via OSM is an easy and cost effective solution; the main problem lies in the effective processing of the data, as explained in Section 3. The images acquired via OSM are typically noisy and may contain surfaces or regions other than lakes. We have developed a framework for detecting HABs from images collected by lay citizens to complement traditional approaches, such as *in-situ* sampling and satellite remote sensing, which also suffer from critical shortcomings as explained in Section 1. We propose a *two-stage* classification framework to first extract *lake* regions in the image using agglomerative clustering-based optimization, and then classify the extracted *lake* regions as containing HABs or not. The current framework is the first step in the design of a comprehensive automated system for early detection of CyanoHAB contamination. Our ultimate goal is to design a unified framework for robust detection of CyanoHABs in their early stages of development by integrating multimodal data such as citizen-science image data and satellite image data, *in situ* sensor data, and high-level textual information obtained via analysis of *tweets* and OSM posts (by leveraging recent advances in Natural Language Processing).

5. Dataset

To evaluate the performance of our approach, we propose a new benchmark dataset consisting of images of natural scenes of inland water bodies (with and without HABs), obtained using our mobile citizen science app, as well as images gathered from other sources such as *Google Images* and *Flickr*. Figure 2 shows a few sample images from our dataset. Our dataset currently consists of 316 images, where we randomly assign 200 images to the training set and 116 images to the test set. We have identified 5 surface categories that occur predominantly and consistently in all images across our dataset, i.e., *lake (clear), tree, grass, sky* and *lake (HAB)*. We have also identified a few more categories such as *mountain, sand,* etc.; but since we currently do not have sufficient training or test instances for these categories, we label them as a single background category. The general paucity of images in our dataset can be attributed to the fact that we are dealing with a very specific application for which there are very few publicly available benchmark images; hence our focus on acquisition and analysis of crowd-sourced images. Owing to the nature of the application, the crowd-sourced images typically exhibit a significant class imbalance. For example, the images containing lakes with HABs are significantly fewer than those containing clear lakes. But, since our primary goal is to detect

HAB-infested lake regions, we aim to obtain as many images of HAB-infested lakes as possible; images of clear lake regions can be efficiently used as a negative set for hard mining of negative instances. We have also developed an annotation toolbox that allows users to manually segment different regions of an image by drawing boundaries around them and to identify and annotate (i.e., tag) these regions. The toolbox also allows users to report noisy images, i.e., images that are not of natural scenes or ones that do not contain lake regions.

6. Agglomerative clustering-based optimization

The goal of the optimization step in the proposed framework is to group superpixels so as to maximize the likelihood of a heterogeneous or cluttered surface category. The proposed optimization procedure effectively integrates the extracted textural and color cues with the learned contextual cues, by iteratively clustering the super-segments or superpixels using a convergence criterion based on a pixel-level *entropy* function [36]. The purpose of the proposed optimization procedure is to arrive at an optimal labeling of each superpixel with a label of one of the predefined surface categories. The proposed agglomerative clustering combines multiple cues (i.e., color, textural, and contextual) while minimizing a pixel-level entropy function via grouping of superpixels so as to maximize the appearance of a heterogeneous surface category. The agglomerative clustering procedure guided by the aforementioned convergence criterion eliminates the need for deciding the final number of clusters before hand.

An individual superpixel $S_i \in S$, provides only a *local* view to a region-based classifier, which is typically insufficient to ensure *global* agreement with its surface category, especially when dealing with surface categories that contain heterogeneous textural patterns. The goal of the agglomerative clustering procedure is to iteratively group the superpixels to reduce the *overall* entropy and thereby obtain *globally* meaningful regions. The optimal segmentation G^* is given by the minimization of the pixel-level entropy function as follows:

$$G^* = \arg\min_{G} \quad U(G) \tag{1}$$

where

$$U(G) = \sum_{l=1}^{N} \sum_{j=1}^{|S|} -P(l|S_j) \log(P(l|S_j)|S_j| \tag{2}$$

Here $U(G)$ is the *pixel-level entropy* of the segmentation state G, $|S_i|$ denotes the total number of pixels in the superpixel S_i , $|S|$ denotes the total number of superpixels in the test image I, and N denotes the number of surface categories.

$P(l|S_i)$ represents the probability of the segment, superpixel or region S_i belonging to a class $l \in N$ and is given by:

$$P(l|S_i) = w^{(1)} P(l|A_{S_i}) + w^{(2)} P(l|D_{S_i}) \tag{3}$$

where $0 \leq \{w^{(1)}, w^{(2)}\} \leq 1$ and $w^{(1)} + w^{(2)} = 1$. $P(l|A_{S_i})$ is the probability of a region or superpixel S_i belonging to the class l, given its appearance score A_{S_i}. We use Platt's scaling [28] to convert Support Vector Machine (SVM) classification scores to probabilities. Similarly $P(l|D_{S_i})$ represents the probability of the superpixel S_i belonging to class l given its 2D location, as explained in Section 6.1.3. Weights $\{w^{(1)}, w^{(2)}\}$ are learned by maximizing the classification likelihood on the training data.

For an image *I*, we initially compute a set of superpixels $\{S_i | i = 1 \dots K\}$ using the method discussed in Section 6.2.1. Subsequently, for each superpixel S_i, the probability that the superpixel belongs to class $l \in L$, is computed using equation (3). A pair of superpixels (S_i, S_j) can be merged if (a) the superpixels are adjacent to each other, and (b) the merging reduces the overall pixel-level entropy. The impact of merging a pair of superpixels (S_i, S_j) on the overall pixel-level entropy is denoted by $A(S_i, S_j)$ and is given by:

$$A(S_i, S_j) = \begin{cases} R(S_i, S_j) & \text{if}, S_j \in \mathcal{N}_{S_i} \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

where $\mathcal{N}_{S_i}$ is the adjacency neighborhood of S_i. $R(S_i, S_j)$ is computed as follows:

$$R(S_i, S_j) = U(G) - U(G \cup \{S_k\} \backslash \{S_i, S_j\})(|S_i| + |S_j|) \tag{5}$$

where $S_k = S_i \cup S_j$. From equations (2), (3) and (5),

$$\begin{aligned} R(S_i, S_j) = & \sum_{l=1}^{N} P(l|S_k) \log(P(l|S_k)|S_i + S_j| \\ & - \left\{ \left(\sum_{l=1}^{N} P(l|S_i) \log(P(l|S_i)|S_i| \right) \right. \\ & \left. + \left(\sum_{l=1}^{N} P(l|S_j) \log(P(l|S_j)|S_j| \right) \right\} \end{aligned} \tag{6}$$

In order to perform agglomerative clustering of the superpixels, we randomly initialize a set of seed superpixels $S_{init} \subset S$. For each superpixel in S_{init}, we merge it with each of its neighbors and choose the pair that results in the maximum reduction in the overall pixel-level entropy. Also we perform multiple random initializations of S_{init} to avoid convergence to a local minimum.

Figure 2. Sample images from the proposed benchmark dataset.

6.1. Training

6.1.1 Textural Features

We use FV-CNN features [4] to characterize the texture of each image segment. The FV-CNN features are an orderless representation of the CNN features and hence ideally suited for texture recognition. FV-CNN features are shown to outperform most state-of-the-art texture detectors [4]. We improvise the approach in [4] by removing the FC-CNN features from their architecture. Since the FC-CNN features provide high-level shape information, they were deemed redundant for our purpose. The FV-CNN is obtained by computing the FV representation over the CNN features. Also, for practical purposes, we extract FV-CNN features for the entire image once, and then compute the scores and histograms individually for each super-segment. This allows efficient search for combinations of superpixels that maximize the appearance of heterogeneous surfaces. To encode the extracted features, we compute η (where $\eta = 64$) Gaussian priors from the data for robust representation of the underlying feature distribution, and using the priors we embed the data into an FV.

6.1.2 Color Features

Instead of using the traditional linguistic label-based representation of colors, we extend the *discriminative color descriptors* (DCDs) [13] approach that clusters color values based on their discriminative power in a classification task. We use the DCDs extracted using [13] and compute a FV representation similar to the one for textural features described in Section 6.1.1. We compute η (where $\eta = 32$) Gaussian priors from the data for embedding the data into an FV representation.

6.1.3 Location Context

Since surfaces in natural scenes are cluttered and sometimes textureless, the surface classification results could be potentially ambiguous. To address this problem, we learn contextual information about the surfaces and use it to resolve the ambiguity. *Location context* denotes the information about the location of a surface with respect to other surfaces that can be used as a prior to improve classification performance. Location context is easy to learn and has been shown to perform well in object recognition tasks. In our case, we learn the location context in order to predict the likelihood of a superpixel belonging to a specific surface category. To learn the location context, we simply average the occurrences of each surface category at each location. For each surface category, we load all the annotated instances and compute binary instance masks. By projecting the instance masks into a 1×1 2D unit space, (which can be denoted as the *standard image*) we first compute the likelihood of occurrence of each surface category at each point in the standard image space. We then compute the joint probability of occurrence for each surface category, using the computed individual likelihoods. In addition, we also compute the pairwise (conditional) probabilities, i.e., given that a surface category occurs at a certain location in the standard image space, we compute the likelihood of occurrence for other surface categories at each point in the standard image space.

We used 200 images from the training set to learn the location context. Figure 3 is a visualization of the individual and joint probability of occurrences for each surface category, learned from the training set using the method described above. As shown in Figure 3, surface categories such as *lake* or *tree*, appear more uniformly distributed since they occur in almost every image in our dataset. In contrast, categories such as *mountain* or *grass* have a patchy appearance in Figure 3 since they occur much less frequently.

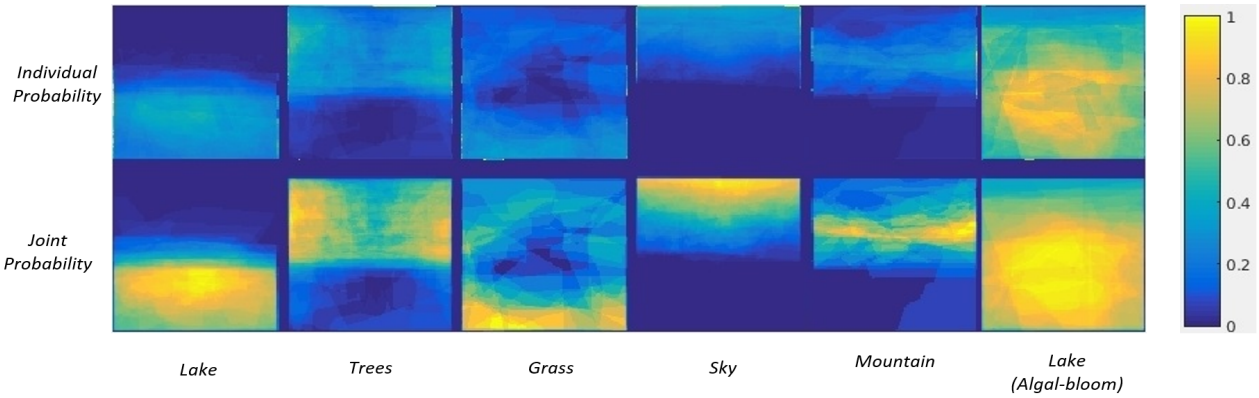

Figure 3. Learned location context: The *first* row represents the joint probabilities and the *second* row represents the individual probability of occurrence of each surface category. *Note:* For better visualization, the joint probability values in the second row are normalized to 1.

Algorithm 1 Agglomerative Clustering

1: Randomly initialize a set of seed superpixels $S_{init} \subset S$;
2: Compute pixel-level entropy $U(G)$;
3: **do**
4: Assign $\hat{U}(G) = U(G)$
5: **for each** superpixel $S_i \in S$, **do**
- Compute $P(l|S_i)$ using equation (3).
- Compute $A(S_i, S_j)$ for each neighbor S_j using equation (4);
6: Merge the superpixel pair with the least entropy.
7: Compute combined pixel-level entropy $U(G)$,
8: **while** $U(G) < \hat{U}(G)$

6.2. Testing

6.2.1 Image segmentation using superpixels

To generate superpixels, we use the simple linear iterative clustering (SLIC) procedure [1] to over-segment the image. In our experiments, we use the SLIC procedure with smaller region sizes and a higher value for the regularizer, so that the resulting segmentation aids the iterative clustering procedure during optimization.

6.2.2 Agglomerative clustering

We perform surface categorization for each superpixel $S_i \in S$. A set of trained SVM-based appearance classifiers followed by Platt's scaling is used to obtain the probability values for each positive response. The agglomerative clustering procedure is outlined in Algorithm 1. In order to prevent the optimization from converging to a local minimum, we perform multiple initializations of the agglomerative clustering procedure by randomly generating different initial sets of seed superpixels $S_{init} \subset S$ (Algorithm 1).

6.2.3 Detection of HABs

Once we extract the pixels that correspond to *lake* regions, we deploy an instance-level *binary* SVM classifier trained to further classify the *lake* regions as *HAB* vs. *clear* lake surfaces. We train the SVM classifier on features extracted from *HAB* and *clear* lake images.

7. Experimental results

We demonstrate the performance of the proposed texture recognition technique which is modeled as a multi-class classification problem followed by a fine-grained classification problem. We first extract *lake* regions from the test image followed by fine-grained classification of the extracted *lake* region to determine whether it contains an HAB or not. To the best of our knowledge, most existing approaches rely on either satellite images or *in situ* water sampling to detect HABs in water bodies; there are very few full-fledged image-based approaches to detect HABs [16].

7.1. Agglomerative clustering-based optimization

Results of the proposed agglomerative clustering-based optimization procedure (i.e., *Stage 1* classification) for joint classification and segmentation of multiple image surfaces are shown in Table 1. The results demonstrate the usefulness of *location context* in reducing ambiguity and enhancing classification accuracy, as well as the improvement in classification accuracy due to the agglomerative clustering-based optimization procedure. Figure 4 shows qualitative results for joint classification and segmentation of multiple image surfaces. Similarly, Figure 5 demonstrates the qualitative improvement in classification accuracy due to the inclusion of location context.

7.2. Classification of lakes: HAB vs. clear

Results of the proposed scheme for binary classification of *lake* regions into *clear* and *HAB* categories based on tex-

Table 1. Performance analysis of the proposed system for joint classification and segmentation of image surfaces using precision and recall measures. The performance of the proposed system is compared to the performance of the scheme described in [36] which does not use *agglomerative clustering* and the performance of proposed system without the use of *location context*.

Categories	Proposed System		w/o. Agg. Clustering [36]		w/o Location Context	
	Precision	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*
Lake (clear)	**0.78**	**0.61**	0.54	0.52	0.74	0.58
Tree	**0.71**	**0.58**	0.60	0.44	0.67	0.51
Grass	**0.68**	0.71	0.57	0.58	0.66	**0.73**
Sky	**0.84**	**0.88**	0.72	0.66	0.77	0.80
Lake (HAB)	**0.71**	**0.68**	0.61	0.46	0.64	0.48

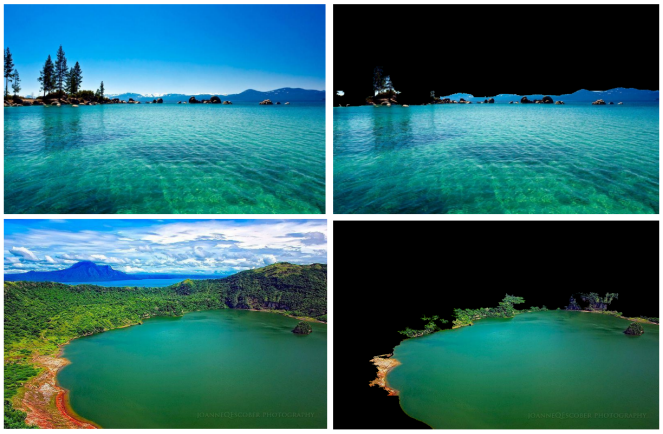

Figure 4. Qualitative results of the proposed agglomerative clustering-based optimization for joint detection and segmentation of clear and HAB *Lake* regions. The *left* column shows test images, where the *right* column shows extracted lake regions.

Figure 5. Improvement in classification accuracy due to location context. *Left:* Test image; *Middle:* Extracted *lake* region without location context; *Right:* Improvement in classification using location context.

tural (FV-CNN) and color (DCD) features (i.e., *Stage 2* classification) are given in Table 2. The proposed scheme is compared with the scheme of Lazorchak et al. [16] which performs binary classification of lake regions into *clear* and *HAB* categories using color histogram features. We have implemented the scheme of Lazorchak et al. [16] using hue-saturation-intensity (HSV) histograms with 300 bins and a binary SVM classifier. Note that the scheme of Lazorchak et al. [16] is restricted to binary classification and, unlike the proposed framework, does not perform segmentation of multi-surface (i.e., heterogeneous) images. Table 2 shows that, given the segmented *lake* regions, the proposed scheme performs significantly better than the scheme of Lazorchak et al. [16] in classifying lake regions as either *clear* or *HAB* surfaces.

Table 2. Comparison of HAB detection performance within *lake* regions (using precision and recall) of the proposed approach and that of Lazorchak et al. [16].

	Proposed Scheme	Lazorchak et al. [16]
Precision	**0.87**	0.764
Recall	**0.844**	0.81

8. Conclusions and Future Work

We have demonstrated quantitatively that the proposed *agglomerative clustering*-based approach for automatic detection of HABs using images acquired by lay citizens, can yield good results, providing an effective and cost-efficient means for monitoring of HABs in inland water bodies. The current system is an initial step towards a design of an automated early detection, warning and rapid response system, to mitigate the detrimental effects of CyanoHAB contamination. In future, we intend to integrate multi-modal information, such as citizen-science image data, satellite image data, *in situ* sensor data and textual information obtained via tweets and OSM posts within the current framework. Our ultimate goal is to design a unified framework for robust detection of CyanoHABs in their early stages of development.

Acknowledgment: This research is supported by the National Science Foundation (NSF) under Grant No. 1442672. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

[1] Achanta, R. et al. (2010) SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. PAMI*, Vol. 34(11), pp. 2274-2282. 7

[2] Backer, L.C. (2002). Cyanobacterial harmful algal blooms: developing a public health response, *Lake Resvr. Mgmt.*, Vol. 18(1), pp. 20-31. 1

[3] Brezonik, P.L. et al. (2015). Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters, *Remote Sens. Envrnmnt.*, Vol. 157, pp. 199-215. 2

[4] Cimpoi, M. et al. (2015) Deep filter banks for texture recognition and segmentation, *Proc. IEEE Conf. CVPR*, pp. 3828-3836. 4, 6

[5] Cimpoi, M. et al. (2015). Describing textures in the wild, *Proc. IEEE Conf. CVPR*, pp. 3606-3613. 3

[6] Codd, G.A. et al. (1999). Cyanobacterial toxins, exposure routes and human health, *Eur. Jour. Phycol.*, Vol. 34(4), pp. 405-415. 1

[7] Dekker A.G. & H.J. Hoogenboom (1997). *Operational tools for remote sensing of water quality: A prototype toolkit.* The Netherlands Remote Sensing Board (BCRS) Report 96-18, Vrije Universiteit Amsterdam, Institute for Environmental Studies. 2

[8] Divvala, S.K. et al. (2009). An empirical study of context in object detection. *Proc. IEEE Conf. CVPR*, pp. 1271-1278. 4

[9] Falconer, I.R. & A.R. Humpage (2005). Health risk assessment of cyanobacterial (bluegreen algal) toxins in drinking water, *Intl. Jour. Environ. Res. Public Health*, 2(1), pp. 43-50. 1

[10] Khan, F. S. et al. (2012) Color attributes for object detection, *Proc. IEEE Conf. CVPR*, pp. 3306-3313. 3

[11] Khan, F.S. et al. (2012). Modulating shape features by color attention for object recognition, *Intl. Jour. Computer Vision*, Vol. 98(1), pp. 49-64. 3

[12] Khan, F.S. et al. (2013). Evaluating the impact of color on texture recognition, *Proc. Intl. Conf. CAIP*, Vol. 8047, pp. 154-162. 3, 4

[13] Khan, R. et al. (2013). Discriminative color descriptors, *Proc. IEEE Conf. CVPR*, pp. 2866-2873. 6

[14] Khan, F.S. et. al. (2015). Compact color/texture description for texture classification, *Patt. Recog. Letters*, Vol. 51(1), pp. 16-22. 3, 4

[15] Krizhevsky, A. et al. (2012). Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.*, Vol. 1, pp. 1097-1105. 3

[16] Lazorchak, J. et al. (2016). Harmful algal bloom smart device application: using image analysis and machine learning techniques for early classification of harmful algal blooms, *Proc. SETAC Europe 2016*. 7, 8

[17] Leung, T. & J. Malik (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons, *Intl. Jour. Computer Vision*, Vol. 43(1), pp. 29-44. 3

[18] Li, L-H. et al. (2015). Remote sensing of freshwater cyanobacteria: An extended IOP Inversion Model of Inland Waters (IIMIW) for partitioning absorption coefficient and estimating phycocyanin, *Remote Sens. Envrnmnt.*, Vol. 157, pp. 9-23. 2

[19] Lunetta, R.S. et al. (2015). Evaluation of cyanobacterial cell count derived from MERIS imagery across the eastern USA, *Remote Sens. Envrnmnt.*, Vol. 157, pp. 24-34. 2

[20] Matthews, M.W. et al. (2012). An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters, *Remote Sens. Envrnmnt.*, Vol. 124, pp. 637-652. 2

[21] Matthews, M.W. (2014). Eutrophication and cyanobacterial blooms in South African inland waters: 10 years of MERIS observations, *Remote Sens. Envrnmnt.*, Vol. 155, pp. 161-177.

[22] Matthews, M.W. & S. Bernard (2015). Eutrophication and cyanobacteria in South Africa's standing water bodies: A view from space, *South African Jour. Sci.*, Vol. 111, Nos. 5/6, May/June, pp. 1-8.

[23] Matthews, M.W. & D. Odermatt (2015). Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters, *Remote Sens. Envrnmnt.*, Vol. 156, pp. 374-382. 2

[24] Mishra, S. et al. (2009). A novel model for predicting phycocyanin concentrations in lab cultured cyanobacteria: A proximal hyperspectral remote sensing approach, *Intl. Jour. Remote Sensing*, Vol. 1, pp. 758-775. 3

[25] Mishra S. et al. (2014). Bio-optical inversion in highly turbid and cyanobacteria dominated waters, *IEEE Trans. Geosci. Remote Sensing*, Vol. 52(1), pp. 375-388.

[26] Ogashawara, I. et al. (2013). A performance review of reflectance based algorithms for predicting phycocyanin concentrations in inland waters, *Remote Sensing*, Vol. 5(10), pp. 4774-4798. 3

[27] Oyama, Y. et al. (2015) Distinguishing surface cyanobacterial blooms and aquatic macrophytes using Landsat/TM and ETM+shortwave infrared bands, *Remote Sens. Envrnmnt.*, Vol. 157, pp. 35-47. 2

[28] Platt, John. et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, Vol. 10(3), pp. 61-74. 5

[29] Quattoni, A. & A. Torralba (2009). Recognizing indoor scenes, *Proc. IEEE Conf. CVPR*, pp. 413-420. 3

[30] Ren, X-F. & J. Malik (2003). Learning a classification model for segmentation, *Proc. IEEE ICCV*, Vol. 1, pp. 10-17. 3

[31] Russakovsky, O. et al. (2015). ImageNet large scale visual recognition challenge, *Intl. Jour. Computer Vision*, Vol. 115(3), pp. 211-252. 3

[32] Schalles, J. & Y. Yacobi (2000). Remote detection and seasonal patterns of phycocyanin, carotenoid and chlorophyll-a pigments in eutrophic waters, *Hydrobiol. Sp. Issues Adv. Limnol.*, Vol. 55, pp. 153-168. 2

[33] Sharan, L. et al. (2013). Recognizing materials using perceptually inspired features, *Intl. Jour. Computer Vision*, Vol. 103(3), pp. 348-371. 3

[34] Sharif Razavian, A. et al. (2014). CNN features off-the-shelf: An astounding baseline for recognition, *Proc. IEEE Conf. CVPR Workshops*, pp. 806-813. 3

[35] Simis, S.G.H. et al. (2005). Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water, *Limnol. Oceanogr.*, Vol. 50(1), pp. 237-245. 2

[36] Vijayanarasimhan, S. & K. Grauman (2010) Top-down pairwise potentials for piecing together multi-class segmentation puzzles. *Proc. IEEE CVPR Workshops*, pp. 25-32. 3, 4, 5, 8

[37] Wilde, S.B. et al. (2005). Avian vacuolar myelinopathy (AVM) linked to exotic aquatic plants and a novel cyanobacterial species, *Environ. Toxicol.*, Vol. 20, pp. 348-353. 1