Image Analysis for Identifying Mosquito Breeding Grounds

Maanit Mehra, Aditya Bagri, Xiaofan Jiang
Department of Electrical Engineering
Columbia University
New York, NY 10027 USA
maanit.mehra, aditya.bagri@columbia.edu,
jiang@ee.columbia.edu

Jorge Ortiz
IBM T.J. Watson Research Lab
Yorktown Height, NY 10598 USA
jjortiz@us.ibm.com

***Abstract*—In this paper we present a technique for identifying the presence of stagnant water bodies in images taken in various settings. Stagnant water bodies, such as puddles, can become sites for mosquitos to grow, which increases the likelihood of the spread of diseases, such as Zika [1]. We observe that existing techniques perform poorly on images of variable quality, variable backgrounds, and focus. We present a technique that is able to successfully identify puddles in these images with about 88% accuracy and over 95% recall. Our technique is robust to image focus, background, and quality, making it a good choice for datasets produced by various kinds of cameras, including those on mobile phones, thermal imagers and unmanned aerial vehicles (UAVs).**

Keywords*—*Stagnant Water Detection, Zika, Computer Vision, Mosquito-borne diseases, Thermal Image Processing.

I. Introduction

With the recent spread of the Zika virus [1] throughout many parts of the world, researchers have started examining ways to identify potential breeding grounds for the disease. Many diseases, including Zika, spread through mosquitoes. Mosquitoes grow and breed in and around stagnant water puddles. When mosquitoes are known to carry certain diseases, these breeding sites become particularly important to identify. In the case of Zika, the problem is especially pernicious, since the disease is especially dangerous and has spread near urban centers. Identifying puddles throughout a city is a major challenge for officials, as puddles can develop almost anywhere and it is difficult to collect the data necessary to narrow the search to the most likely sites. Moreover, identification of puddles in images is mostly done via 'manual' (visual) inspection. A process that does not scale and makes it difficult to keep up with the spread of diseases. A different approach must be explored.

For example, consider Rio de Janeiro in Brazil, where Zika has been widespread. Breeding sites have been found throughout the city, near roads, in people's homes, in wooded areas, and garbage sites. Especially affected on the shanty towns which contain the kinds of sites that are difficult to find and access [2]. This highlights the challenge with identifying puddle sites from images. Urban settings have non-uniform scenes. In cities like Rio De Janeiro, some images contain urban backgrounds, such as images with concrete and buildings, while other can be filled with shrubs or garbage. Also, many of the places where stagnant puddles form are in poorly lit, poorly ventilated areas, in backyard dens and storage areas. This makes data collected by any single means very difficult. The use of any single modality will affect the types of images that can be taken and miss an entire class of images that can contain breeding sites. For instance, if images are taken from the road or in the air with drones, these will only capture open spaces, where vehicles can drive and drones are allowed the fly.

The data-collection apparatus should not limit ability to detect certain types of sites. A complete solution must analyze images from any modality – including drones, mobile phones, personal cameras, etc. Crowd-sourcing the data collection process would enable us to collect from a wide range of a people, at a large spatio-temporal scale. This approach allows any user to participate and does not hinder organizations from collecting images with their apparatus of choice. However, this also makes it difficult to keep up with the amount of data that is generated. Two solutions are possible: 1) crowd-sourcing the labeling process or 2) automating it. In this paper, we choose the second option because it is less costly than the first one. To automate the inspection process, we must construct a robust computational pipeline, such that variance in picture quality, lighting, scene color mixtures and other factors do not affect it. Specifically, the technique should be able to detect whether or not the image has a puddle, and do so in a broad set of images.

There are many existing techniques that look at the detection of water in images. These works tend to be highly specialized for specific tasks, using images in very narrow domains. For examples, there are algorithms that can detect whether there is a body of water in a satellite image [3]. There are also algorithms that can estimate the depth of that body of water. In our data the color gradients can be much subtle, making it difficult to detect sharp contrasts between the background on the water. Moreover, depth estimation is outside the scope of our work. There are devices that can help detect water presence in dark space using thermal imaging. These devices do not themselves flag an image automatically, instead they show the image to an expert who makes a decision about whether or not they perceive a puddle of water. We leverage those images as well in our work.

In this paper, we propose such a technique. Our approach constructs a pipeline of binary classifiers used as an ensemble to maximize robustness. Moreover, we use boosting to maximize the performance of the ensemble. We are able to achieve about 90% accuracy on a broad set of images taken with different kinds of cameras. We argue that our technique could outperform the state-of-the-art in water identification techniques under a wide range of conditions. We also state that this robustness property is fundamental to achieving scalability in the identification processes. It allows us to incorporate images taken from a wide array of devices and scene conditions, including mobile phones and drones.

A. *Problem Statement*

The crux of the problem lies in determining stagnant water in images, in the form of puddles or other water formations, and calculating the probability of the region depicted in the image being a potential breeding ground of Zika Virus carrying mosquitoes by running the image through a series of trained algorithms. In this project, the results have been compiled on a data set of visible range RGB and thermal images considering one set at a time and ultimately generating results on the combined data set. Post image processing on the two sets of image data, we developed a web application that displays an interactive visualization of a heat map along with numerous sample images. On clicking one such image, a message pop up informs the user whether or not that particular region is a breeding ground and if it is, the region is illuminated red in the heat map visualization.

II. Related Work

There are a number of related works in the identification of water puddles in images taken with various kinds of images, including standard RGB, thermal, and others. The authors of [4] demonstrate a technique that uses infrared thermography to detect fluid leakage. The technique dates back over 20 years and it very reliable. We propose a similar approach except that we train on infrared images to improve the overall classification robustness. It is advantageous to leverage both visual-spectrum and infrared images for water identification. We show that this combination increases our overall accuracy in classification accuracy.

In [5], [6] the authors present three algorithms used in concert to detect water in images for a robotic unmanned ground vehicle. They use mounted stereo cameras to multicue water detector that uses a rule base to combine water cues from color, texture, and object reflections (detectable in stereo range data). Their combination of techniques can achieve nearly 90% accuracy, depending on various factors, such as distance from the body of water. We do not use stereo cameras in our work and their technique is complimentary to ours. We propose a more general solution design for images that come from a wide array of cameras and use infrared thermography, when available. The latter can be mounted on drones, for example, and only serve to improve the results as an extra source of information.

The authors of [7] use a set of wireless sensors and IR cameras to detect water leakage in pipes. Their overall approach is quite different from ours, we focus specifically on evaluating images taken from different types of cameras.

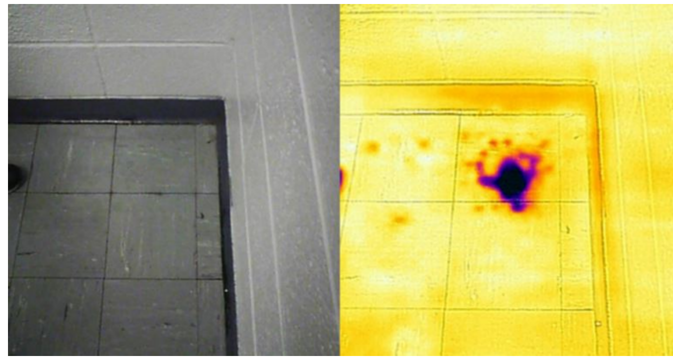

Fig. 1. Distinct advantage of using thermal images to detect the presence of water puddles over standard RGB images

Their work is focused on the wireless sensors and the network their form to collect and report the data collected by embedded infrared cameras. They do not take visual images pipeline nor do they focus on the machine learning algorithms and techniques that may automate the process. We focus on the machine learning steps and techniques to enable robust identification of water in images. We are specifically concerned with breadth and use thermal images to improve the number and types of setting in which we could detect such water bodies. That said, their work does demonstrate that using non-visual images (I)R can be effective at locating water in poorly lit areas.

Our work expands on the work presented by Agarwal et al. [8]. However, there are some very important differences. We do not use K-Means to construct our vocabulary of classifiers [9]. Instead we compress all the features vectors into a single representative vector. We show that this works well for our application use-case. In addition, we reduce the feature vector size using PCA, as described by Valenzuela et al [10] and use SURF [11] rather than SIFT [12] or ORB [13]. Finally, we use an ensemble of Bayesian classifiers [14] instead of SVM [15] and we boost our classifier with Adaboost [16]. We implement a similar version of what they propose and show that we outperform their approach, with a classification accuracy up to 90% versus their 83%.

III. Methodology

This paper focuses on the problem of stagnant water detection. We address this problem with a proposed approach that allows for scalability of systems and a modular approach that allows for the presence of multiple possible input types. We begin by addressing two separate challenges, first: the task of working with data from the ”visible” range; second: proceeding towards the usage of thermal image data alone; third: using a combination of information from both datasets. As described previously, the usage of the visible range allows for detection in the visible range, a wider source of image data and a fairly wide problem set. We address the algorithm below.

A. *Our Dataset*

In order to test and compare approaches, we collect images from various sources; including Google image search and our own cameras. The images vary in quality and size. Some of

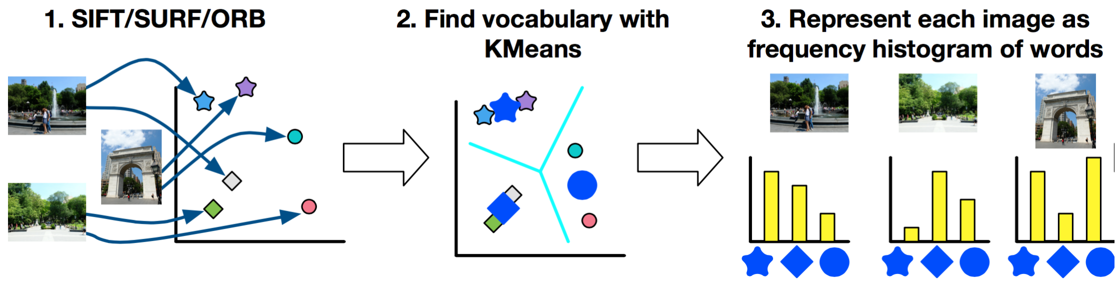

Fig. 2. Canonical bag-of-words feature extraction pipeline. The first step extracts features from the image, the second step clusters the image vectors, and the third step uses the cluster heads as representative images. Our pipeline differs slightly from the canonical approach. We extract features using SURF, reduce the dimensionality of each feature using PCA, and compress the N features into 1.

them have water in the image and others do not. We refer to each unique image as a scene. For some scenes in our dataset, we take both a "regular" picture and a thermal one. For thermal images, we use the FLIR One Thermal Imaging System [17]. Figure 1 shows an example of two images of the same scene. The picture on the left is an RGB image while the one on the right is a thermal image. Note, the puddle cannot be detected in the RGB image while it is quite apparent in the thermal one. Thermal images can improve overall classification accuracy in cases where puddles lie in poorly lit areas and/or not easily captured in the visual spectrum, such as in figure 1.

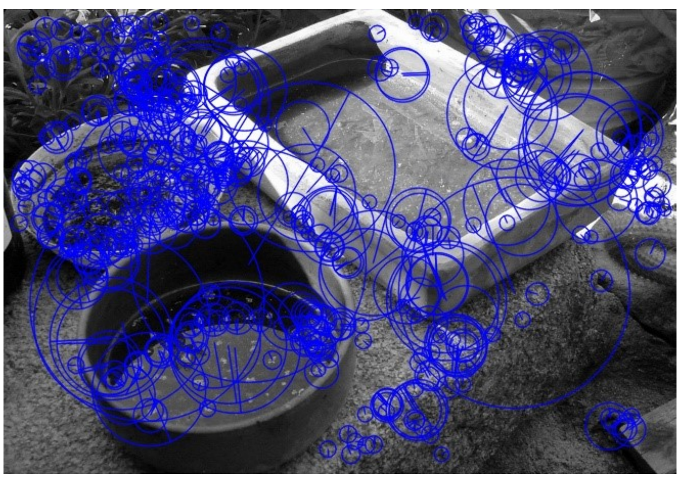

Fig. 3. Result of SURF computation. Extracted features using SURF described by the blue lines in the image

For scenes with puddles, we take either naturally-formed puddle images or "synthetic" ones (water poured on floor). After a rain event, we collect a number of images of puddles throughout our campus. To construct a more realistic dataset, some of the scene contain puddles near dirt patches and/or garbage. Stagnant puddles in scenes like these are the most likely to develop into mosquito breeding grounds. Although our technique does not currently differentiate between stagnant and clear puddles, we believe that having these in our dataset will allow us to experiment with expanding our current puddle-identification technique to automate that differentiation as well.

We collect images from multiple sources and generate transformations on them in order to improve the learner. We collect a number of images from phones and Google image search. We also collect a number of images using our FLIR

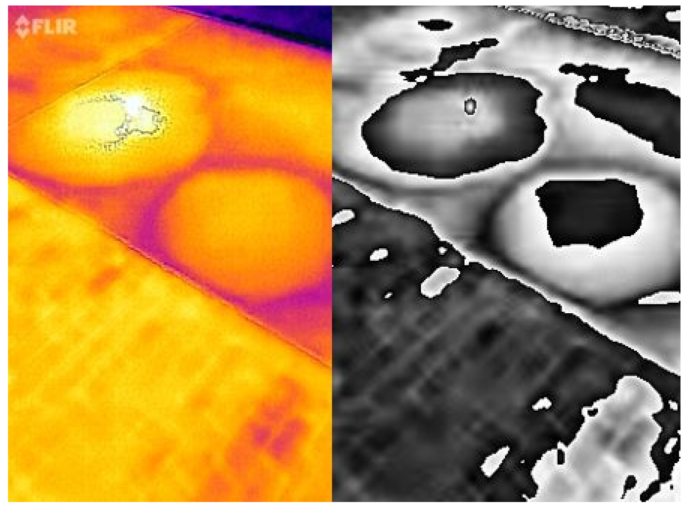

Fig. 4. Thermal Extraction distinctly showing presence of a puddle. The extracted image clearly shows pronounced outlines of water puddles which are not visible in raw thermal image

camera. The FLIR camera generates both RGB and thermal images for a scene, combined together into a single object. We separate the two and then generate corresponding grey-scale images for both the thermal image and the RGB image. Although no direct gray-scale image exists for the thermal image one extract one of the color bands from it and effectively use it as the gray-scale analog. An example of the output 'gray-scale' image for the thermal image output is down in Figure 4. In addition, each of these images is then rotated 90, 180, and 270 degrees. We end up with 8 images per scene. In our pre-processing step, however, some of the images became corrupted when a rotation was performed and had to be discarded. We did not note which ones we discarded. It did not have a major impact on our final result.

Our dataset contains 2405 images This includes RGB images, thermal images, and rotated images of the unique scene. We use each as unique inputs to our classifier. Training on *rotated images* has a statistically significant effect on the overall classification accuracy. This effect is captured in our final results. Not all our scenes are unique and not all our images are rotated or have a corresponding thermal image, however. Of the 2405 inputs, 1393 of them are unique scenes. 670 of the images have an RGB and a corresponding thermal image. 723 of them are only available in RGB. About half our images contain puddles in the scene. We train our classifier using 50 random images from our dataset.

B. Challenges

The initial hurdle to our project experiment stemmed from the absence of an extensive research on this topic in the past. Using images, especially thermal images, for the purpose of identifying mosquito breeding grounds is a challenge which has not been addressed before. Choice and combination of algorithms for the best accuracies came about trial and error. Owing to the fact that there is no standard data set, we have proceeded with building our own unique data set of standard RGB and thermal images to test the algorithms. It was our aim to have a hyper-localized approach to the subject. Furthermore, since the number of external variables vary considerably from image to image we intended to design a robust algorithm that integrates the different types of images.

IV. Processing Pipeline

Our image processing deviates from the canonical pipeline for extracting image features. In the canonical structure, features are extracted and smoothed to construct a set of representative 'words' to describe each image. This allows is used to represent each image as a vector of descriptors, also known as a 'bag-of-words'. Figure 2 shows a visual demonstration of this pipeline. We extract features similar to those shown in left-most part of the figure, however, instead of smoothing with a clustering algorithm we first reduce the dimensionality of the extracted features and compress all the features to a single vector. This works well for our dataset because images that contain water puddles are prominently shown in the image and compression makes the associated weight of those features more pronounced for binary classification.

More specifically, we run SURF on each image, which yields a set of N, 128-bit features vectors that correspond to visual artifacts in the scene. An example of SURF features extracted from an image is shown in Figure 3. We reduce the size of this vector to half the number of principal components, with PCA. Each vector 128-bit vector is casted down to 64 bits and we calculate the average of N vectors into 1. What we end up with a single, 64-bit vector representation of the input image. We find that these step are important for accentuating important features in the underlying image. It improves our results and execution time. However, this approach is quite specific to observation we made about puddle scenes in our data set. The puddles dominate the scene, so compressing the features to a single vector brings our the dominant characteristics of the puddle(s) in the scene. Scenes without puddles do not have the same property and our classifier can successfully differentiate between them.

A. Naive Bayes and Ensemble Learning

For our binary classifier we use a Naive Bayes Classifier (NBC). Naive Bayes [14] is a simple but robust, well-performing classifier that is used in many applications. It is used to classify vectors of discrete-valued features $\mathbf{x} \in \{1, \ldots, K\}^D$, where K is the number of values for each feature and D is the number of features. In our application, K is the length of each feature descriptor extracted from the image and D is the number of features in our word vector. Naive Bayes makes the simplifying assumption that the features are conditionally independent. Therefore, we can calculate the posterior as a product of one-dimensional probability densities as follows:

$$p(x|y=c,\theta) = \prod_{j=1}^{D} p(x_j|y=c,\theta_{jc}) \quad (1)$$

Where c is the class label and θ is the set of parameters of the distributions. In our work, a the classifier is binary (0,1), so we use a Bernoulli distribution prior.

1) Ensemble Learning: Although NBC works quite well for most datasets, we found that it was not robust enough on it's own for ours. In order to increase the consistency of our accuracy measurements, we used an ensemble of NBC's. Ensembles [18] learn how to weigh a collection of classifiers on the input data. Those classifiers that are more reliable get a higher weight in the combination of the ensemble. The net effect increases accuracy and stability of the classifier.

2) Boosting: In order to further improve our results, we added a boosting step using the Adaboost algorithm [16]. Boosting is a process by which weak classifiers concentrate on examples that other classifiers in the ensemble are missing. This allows a subset of learners to becomes really good at classifying inputs that are the most difficult to correctly label. We will show that these steps significantly increased the overall performance of our approach.

V. Results

In this section we examine how our approach compares with the SVM-based approach presented in the literature. We also examine how the added steps in the pipeline improve our results. We implement a version of the algorithm presented in [8]. We do not use BRISQUE [19] scores to filter out images of poor quality. We do not need this step because we directly control the quality of the images during the collection phase. We compare the performance of this approach with our pipeline, with and without boosting. We also compare for specific subsets of our dataset, namely RGB images, thermal images, and a combination of the two. The results are presented in Table I.

We observe that without boosting, the SVM-based approach outperforms the Bayesian esemble classifier. We also note that the ensemble classifier performs better on RGB images than on thermal images. However, when boosting in added, we outperform the SVM-based approach by a similar margin by which it outperforms the ensemble NBC without boosting. When we combine the images, we even better. We are able to achieve an average classification accuracy of 90%.

We train our classifier with 50 randomly chosen images and stop the training process when the In order to train each classifier, we take a sample of 50 images from our data set and run the experiment multiple times. Figure 5 shows the results error of the classifier in the training and testing phase. The x-axis is the number of iteration in the boosting phase. Here we show two lines. The 'Training Error' shows the error on the training set in the given iteration. The 'Testing Error' shows t]

Dataset	SVM	Ensemble- No Boosting	Ensemble- With Boosting
Visual	0.83	0.72	0.88
Thermal	0.82	0.65	0.88
Combined	0.82	0.68	0.90

TABLE I. AVERAGE CLASSIFICATION ACCURACY ON DATA SET.

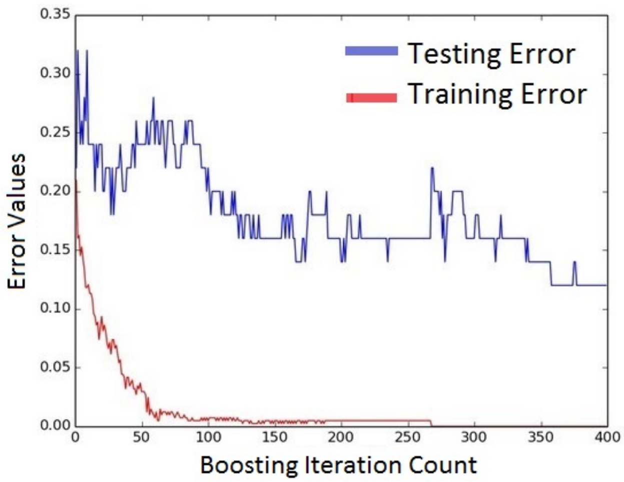

Fig. 5. Training and Testing error for Bayesian Ensemble Classifier. We can see that boosting reduces error by nearly 50%.

the error on the testing set *using the weight calculated in the corresponding iteration of the training set*. The purpose of this graph is the show the effects of boosting on the overall learning results. We observe that boosting reduces classification error by almost 50% after a few hundred iterations. Beyond about 350 iteration, however, we start to overfit.

A. Discussion

The results of our initial experiments are promising. We look to expand upon our approach by building an application that can alert the proper authorities of the possible growth of Zika and similar diseases. Figure 6 shows an example application that we are working on. Using a heatmap visualization, we could highlight and track potential breeding grounds. If made public or used by authorities, it could help both track and contain the spread of dangerous diseases that pose a major health threat.

From a social perspective, community participation continues to be a powerful solution to a problem such as this. In addition to working with the authorities, we hope to include crowd sourcing & labeling as an additional means of data collection. Secondly, while we propose the usage of experimentally-guided feature engineering and classification mechanisms – that currently involves reduced/compressed image data – we believe that using neural networks on raw images could serve to enhance both the quality and performance of the classification mechanism. The technique could be applied equally well to both thermal and visible-range images. We hope to collect sufficient data to be able to use some of these techniques as part of potential future work.

VI. CONCLUSION

The Zika Virus has already become an urgent international concern for health and disease control authorities. Any research to curb the spread of the disease is crucial and valuable. As Brazil aims to host the Olympics this year, Rio de Janeiro is looking for answers to the rampant spread of the Zika Virus. Government authorities and international health organizations including the WHO are increasingly looking to research projects in this space to help inhibit the dangers of the virus. By identifying and mapping mosquito breeding sites on a city level, this project targets a critical contemporary issue and does so without the need for expensive tools, machinery or scientific devices. Images can be collected via publicly installed cameras, Google street view and perhaps even crowd sourcing.

The results of this project proved to be encouraging. We were able to successfully secure a 90% accuracy on the combined visible range RGB and thermal images, and so we can say with considerable precision whether or not the region depicted in the image is a potential breeding ground of the Zika Virus mosquito.

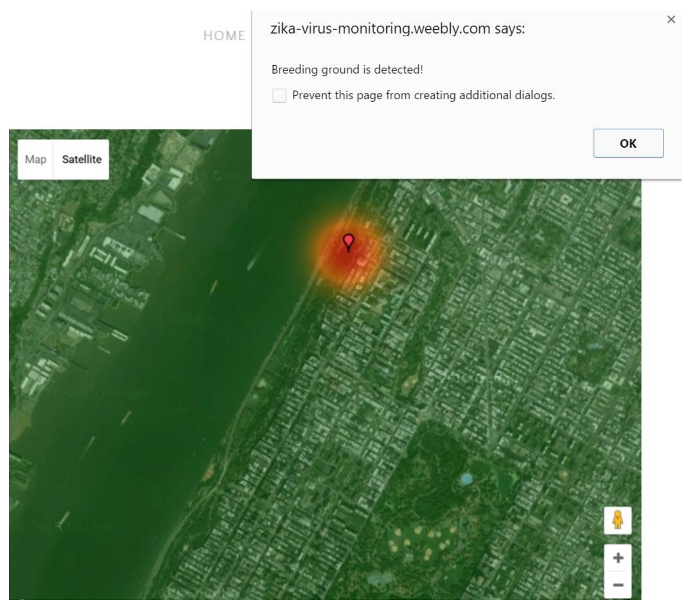

Fig. 6. An example visualization of the resultant heat map showing the presence of a breeding ground and marking the region in its vicinity as a potential breeding site

REFERENCES

[1] E. Petersen, M. E. Wilson, S. Touch, B. McCloskey, P. Mwaba, M. Bates, O. Dar, F. Mattes, M. Kidd, G. Ippolito, E. I. Azhar, and A. Zumla, "Rapid spread of zika virus in the americas - implications for public health preparedness for mass gatherings

at the 2016 brazil olympic games," *International Journal of Infectious Diseases*, vol. 44, pp. 11 – 15, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1201971216000217

[2] "The stinking shanty town exposed as 'gigantic breeding ground' for zika virus," *http://www.dailymail.co.uk/news/article-3436309/The-stinking-shanty-town-exposed-gigantic-breeding-ground-Zika-virus-just-2-5km-Brazil-s-Olympic-airport.html*, 2016.

[3] E. Gedik, U. inar, E. Karaman, Y. Yardmc, U. Halc, K. Pakin, and H. Ergezer, "Automatic water canal detection in multispectral satellite images," in *Signal Processing and Communications Applications Conference (SIU), 2013 21st*, April 2013, pp. 1–4.

[4] G. J. Weil and R. J. Graf, "Infrared-thermography-based pipeline leak detection systems," in *Proc. SPIE*, vol. 1467, 1991, pp. 18–33.

[5] A. Rankin, T. Ivanov, and S. Brennan, "Evaluating the performance of unmanned ground vehicle water detection," in *Proceedings of the 10th Performance Metrics for Intelligent Systems Workshop*, ser. PerMIS '10. New York, NY, USA: ACM, 2010, pp. 305–311. [Online]. Available: http://doi.acm.org/10.1145/2377576.2377632

[6] A. L. Rankin, L. H. Matthies, and A. Huertas, "Daytime water detection by fusing multiple cues for autonomous offroad navigation," in *Proceedings of the 24th Army Science Conference*, Orlando, FL, USA, 2004.

[7] R. K. Parida, V. Thyagarajan, and S. Menon, "A thermal imaging based wireless sensor network for automatic water leakage detection in distribution pipes," in *Electronics, Computing and Communication Technologies (CONECCT), 2013 IEEE International Conference on*, Jan 2013, pp. 1–6.

[8] A. Agarwal, U. Chaudhuri, S. Chaudhuri, and G. Seetharaman, "Detection of potential mosquito breeding sites based on community sourced geotagged images," vol. 9089, 2014, pp. 90 890M–90 890M–8. [Online]. Available: http://dx.doi.org/10.1117/12.2058121

[9] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval*, ser. MIR '07. New York, NY, USA: ACM, 2007, pp. 197–206. [Online]. Available: http://doi.acm.org/10.1145/1290082.1290111

[10] R. E. G. Valenzuela, W. R. Schwartz, and H. Pedrini, "Dimensionality reduction through pca over sift and surf descriptors," in *Cybernetic Intelligent Systems (CIS), 2012 IEEE 11th International Conference on*, Aug 2012, pp. 58–63.

[11] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, Jun. 2008. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2007.09.014

[12] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*, ser. ICCV '99. Washington, DC, USA: IEEE Computer Society, 1999, pp. 1150–. [Online]. Available: http://dl.acm.org/citation.cfm?id=850924.851523

[13] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proceedings of the 2011 International Conference on Computer Vision*, ser. ICCV '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 2564–2571. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2011.6126544

[14] P. Langley, W. Iba, and K. Thompson, "An analysis of bayesian classifiers," in *Aaai*, vol. 90, 1992, pp. 223–228.

[15] I. Steinwart and A. Christmann, *Support Vector Machines*, 1st ed. Springer Publishing Company, Incorporated, 2008.

[16] R. E. Schapire, "Explaining adaboost."

[17] Comfy, http://www.flir.com/flirone/.

[18] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," *Artif. Intell. Rev.*, vol. 11, no. 1-5, Feb. 1997.

[19] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov 2011, pp. 723–727.