# Improved Supervised Classification of Underwater Military Munitions Using Height Features Derived from Optical Imagery

Arthur C. R. Gleason
University of Miami, Coral Gables, FL, USA

Nuno Gracias, ASM Shihavuddin
University of Girona, Girona, Spain

ASM Shihavuddin
Institut de Biologie de l'Ecole Normale Supérieure, Paris, France

Gregory Schultz
White River Technologies, Inc., Lebanon, NH, USA

Brooke E. Gintert
University of Miami / RSMAS, Miami, FL, USA

***Abstract*—A supervised classification routine was used to classify munitions targets and basic seabed types from underwater images. Additional features that were based on the local relief, or height, of the seabed were then added to the classifier and new results computed using the expanded feature set. The height data were generated from the input images themselves using structure-from-motion computer vision techniques. The initial image classifier was shown to distinguish munitions from non-munitions (background) with generally > 80% accuracy except that many false positive matches for munitions were observed. Extending the algorithm to also use height data derived from stereo reconstruction showed that incorporating such "2.5-D" data greatly improved the classification results. Using the 2.5-D information reduced the number of false positives. Furthermore, improved accuracy was observed not only on the basic, binary munitions / non-munitions classes. Adding 2.5-D information also improved the capability to discriminate different types of munitions from one another.**

***Keywords—underwater military munitions, seabed classification, 2.5-D features***

## I. Introduction

The Department of Defense needs cost-effective methods for locating and identifying navigation and safety hazards related to underwater military munitions (UWMM). Technologies are needed that can efficiently and objectively detect, identify, and map underwater military munitions. Furthermore, knowledge of benthic environments adjacent to UWMM is critical for remediation decisions. Currently, both wide-area searches and detailed mapping of underwater military munitions rely on multiple acoustic (*e.g.* side-scan or multi-beam sonar) and/or metal detection methods. Underwater optical images of the seabed are another technology that could benefit surveys for UWMM as well as efforts to understand the environments around UWMM because optical images have very high spatial resolution.

Efforts to automate classification of the seabed using optical imagery have been made for well over a decade [1-5] but no single algorithm is yet widely robust across different habitats and depths. A few of the most challenging obstacles to classification accuracy in natural environments include: significant intra-class and inter-site variability in the morphology of benthic organisms, complex spatial borders between classes on the seabed, subjective annotation of training data by different analysts, variation in viewpoints, distances, and image quality, limits to spatial and spectral resolution when trying to classify to a free taxonomic scale, partial occlusion of objects due to the three-dimensional structure of the seabed, lighting artifacts due to wave focusing, and variable optical properties of the water column.

To overcome some of these challenges, a seabed classification algorithm for optical images was recently developed by our team [6]. The algorithm used a feature vector computed from completed local binary pattern (CLBP), grey level co-occurrence matrix (GLCM), Gabor filter response, and color histograms. Either k-nearest neighbor (KNN), support vector machine (SVM) or probability density weighted mean distance (PDWMD) classifiers were used, depending on the characteristics of the dataset. The algorithm has been tested on datasets from Florida, USA, Moorea, French Polynesia, and the Red Sea with overall accuracy of 96% on 14 classes, 85% on 9 classes, and 97% on 8 classes, respectively [6]. The method used both texture and color, and, based on the datasets previously tested, was a good starting point for discrimination of UWMM.

The goal of this experiment was to test the capability of the algorithm [6] to classify munitions targets in addition to basic seabed types. In addition, we also investigated the utility of adding additional features to the classifier that were based on the local relief, or height, of the seabed. Height data were generated from the input images themselves using structure-from-motion computer vision techniques.

This work was supported by the Strategic Environmental Research and Development Program, project MR2414, and by the ACCIO/TecnioSpring program of the Generalitat de Catalunya (TECSPR14-1-0050).

## II. Materials and Methods

### A. Image Dataset

Underwater images of the seabed and artificial targets were collected in the vicinity of Key Biscayne, FL. Two test sites were used, one at a coral reef and one over seagrass. At each site a suite of test objects were lowered to the seabed then arranged by divers into two parallel rows containing 6 targets each. The targets were separated by ~2 m along each line and the two lines were themselves separated by ~3 m. A variety of inert munitions were used as the test targets (Table I).

TABLE I. Inert Munitions Used as Test Targets

| Abbreviation[a] | *Description* |
|---|---|
| MA | 60mm Mortar (without fuze) |
| MB | 60mm Mortar (with fuze) |
| MC | 81mm M43A1/M49A2 Mortar |
| MD | 81mm Mortar (with fuze) |
| ME | 81mm M821A1/M889A1 |
| MF | 2" Rocket APFSDS-T M735 |
| MG | 76mm Projectile (no fuse) |
| MH | 3" Armor Piercing Projectile MK28 Type A |
| MI | 90mm Projectile |
| MJ | APDS Adapter with End Cap and Cartridge |
| MK | 105mm Projectile (Blue Training Round) |
| ML | 105mm Projectile (with band and solid tip) |

[a.] Abbreviation used in Tables V, VI

Data were collected uing a Nikon D7000 camera with 24-mm lens in an Ikelite underwater housing. The camera was set to shutter speed priority (1/400 second) in order to eliminate motion blur. Therefore, aperture and gain (ISO) varied to accommodate the chosen shutter speed. Images were captured at 1 s intervals as a diver swam over the test targets. The overlap between consecutive images varied based on the diver's speed and 1 Hz frame rate. Consequently, each section of the bottom was captured in 4-6 consecutive frames.

After the diver with the cameras made a pass over the targets, a second diver then moved all of the targets, changing their orientation and location on the seabed. In this way, the parallel lines of targets moved, enabling multiple images of each target in different poses and with different local backgrounds.

A subset of the complete dataset was chosen for processing. Two subsets from each of the reef and seagrass habitats were chosen. Each subset contained a sequence of 25-30 overlapping images, which were processed as described in the following sections. A selection of overlapping images was needed in order to derive the height maps described in Section II.C.

### B. 2-D Image Classification

As mentioned above, four series of images were processed, two from reef and two from seagrass. An existing algorithm [6], which will be called the "2-D algorithm" here for convenience, was used to classify non-overlapping images selected from each series. Non-overlapping images were used because the evaluation step (Secion II.D) assumed independent, random samples, which would not be possible if multiple images viewed the same target. Each of the image series contained 6-8 non-overlapping images.

For each of the images being classified, a set of 100 points were randomly distributed on the image and an analyst identified what was under each point, either one of the 12 classes in Table I or background (*i.e.* natural seabed). Hardly any random points fell on the test targets, however, because as a percentage of the area covered by the dataset the "background" classes dominated. Thus, additional, non-random ground truth points were added by the analyst to ensure that at least 100 points were defined for each class being used. The ground truth points were separated into training and evaluation subsets as described in Section II.D.

### C. 2.5-D Image Classification

The 2-D algorithm [6] was based entirely on image color and texture features. An extension to the 2-D algorithm was developed that used a digital elevation model (DEM) to add features to the classifier based on the height of the seabed [7]. We call the DEM approach a "two and a half dimensional" (2.5-D) method because it only captures information that can be seen from the overhead viewpoint of the camera as opposed to a true three-dimensional (3-D) method, which would capture information even on surfaces that can't be seen from overhead.

As described by [7], hereafter called the "2.5-D algorithm", the first step was to generate a digital elevation map from each sequennce of overlapping images using a structure from motion (SfM) technique. Next, the following 2.5-D features were extracted from each DEM: polynomial surface coefficients [8], elevation statistics [9], slope of surface [10], curvature [11], surface normal [9], rugosity [10] and symmetry [12]. Details of the features were explained in [7]. Finally, the new 2.5-D features were appended to the original 2-D features from [6]. The expanded feature set was the only difference between the 2-D and 2.5-D algorithms, which were otherwise identical.

### D. Evaluation

Classification assessment was conducted using the error matrix approach [13]. The error matrix uses $N$ points within the classified dataset that have known, verified class identification for comparison with the classification results. For the Reef 1, Reef 2, Seagrass 1 and Seagrass 2 datasets $N$ = 371, 368, 389, 365, respectively for the 2-D algorithm and 223, 325, 285, 241 for the 2.5-D algorithm. These points were independent from the ground truth points that were used to train the classifiers.

An error matrix is a table in which the rows represent the classes identified by the ground truth and the columns represent the classes identified on the classified image. Each element in the raw error matrix is a sum of the number of reference points with a partiuclar ground truth class / image class combination. Normalizing the raw error matrix in different ways produces different measures of the quality of the classification. The simplest metric is the overall accuracy (OA), which is the number of reference points correctly classified divided by the total number of reference points. Normalizing the error matrix by the row or column sums reveals two other metrics called "Producers" and "Users" accuracy, respectively [13]. Normalizing by the row sums, (*i.e.* the number of points in each class in the reference dataset), gives an indication of the rate of *false negatives* on a per-class

basis. Conversely, normalizing by the column sums, (*i.e.* the number of points in each class in the classified, or predicted, dataset), gives an indication of the rate of *false positives* on a per-class basis.

The overall, producers, and users accuracy were computed for each of the four image series as processed using both the 2-D and 2.5-D algorithms. In addition, these statistics were computed using the full range of 13 classes, which discriminated the munitions types from one another and the background, as well as a reduced set of classes consisting of only munitions and background. The simple binary classification scheme was of interest because it may often be useful to identify presence/absence of munitions even if automated identification is not possible.

## III. Results

Overall accuracy ranged from 53 to 95% across the four datasets and two methods of processing (Table II; note all values in Tables II-VI are percentages). One pattern observed in the OA data was that accuracy increased as fewer classes were used. OA using a binary classification sheme (munitions vs. background) was never less than OA using the full classification sheme that discriminated munitions types (MTypes; Table II).

A second pattern observed in the OA data was that accuracy of the 2.5-D algorithm was greater than using the 2-D algorithm. Sometimes this difference was substantial, greater than 20 percentage points for many of the image sequences (Table II).

TABLE II. Overall Accuracy

| Dataset | *MTypes* | *Binary* |
|---|---|---|
| Reef 1 2-D | 76 | 80 |
| Reef 1 2.5-D | 94 | 94 |
| Reef 2 2-D | 71 | 75 |
| Reef 2 2.5-D | 95 | 95 |
| Seagrass 1 2-D | 67 | 72 |
| Seagrass 1 2.5-D | 95 | 95 |
| Seagrass 2 2-D | 53 | 60 |
| Seagrass 2 2.5-D | 89 | 89 |

Users and producers accuracy metrics compensate for the limitations of overall accuracy. Namely, OA does not reveal class-by-class information and will be skewed if the reference data were not equally distributed among the classes [13]. Complete users and producers accuracy tables have elements that correspond to every entry in the error matrix. Most of the information in these tables are contained in just the main diagonals, however. Thus, each column in Tables III-VI contained only the main diagonal from a single users or producers accuracy matrix. This format saved space and provided a convenient way to compare results from different sites and with/without the 2.5-D data.

With the binary classification scheme, both users (Table III) and producers (Table IV) accuracy were greater for all four image sequences when using the 2.5-D algorithm than when using the 2-D algorithm. The main errors were false positive matches for munitions. The evidence for this was high producer's accuracy (all munitions were classified as munitions) but lower user's accuracy (many pixels classified as munitions were actually something else).

TABLE III. Users Accuracy: Binary Class Scheme

| | *Reef 1* | | *Reef 2* | | *Seagrass 1* | | *Seagrass 2* | |
|---|---|---|---|---|---|---|---|---|
| **Class** | *2-D* | *2.5-D* | *2-D* | *2.5-D* | *2-D* | *2.5-D* | *2-D* | *2.5-D* |
| Seabed | 94 | 98 | 97 | 98 | 98 | 97 | 98 | 98 |
| Munitions | 25 | 73 | 28 | 76 | 25 | 86 | 18 | 54 |

TABLE IV. Producers Accuracy: Binary Class Scheme

| | *Reef 1* | | *Reef 2* | | *Seagrass 1* | | *Seagrass 2* | |
|---|---|---|---|---|---|---|---|---|
| **Class** | *2-D* | *2.5-D* | *2-D* | *2.5-D* | *2-D* | *2.5-D* | *2-D* | *2.5-D* |
| Seabed | 83 | 94 | 74 | 97 | 70 | 97 | 57 | 89 |
| Munitions | 54 | 91 | 80 | 84 | 86 | 86 | 89 | 90 |

Likewise, when using the full classification scheme that discriminated among the various munitions types, the users (Table V) and producers (Table VI) accuracy were both greater for all four image sequences when using the 2.5-D algorithm than when using the 2-D algorithm. Note that Tables V and VI contain both blank elements and zero values. The zero values were truly zero accuracy, *i.e.* none of the ground truth points were correctly identified for class/datasets with 0 accuracy. The blank entries, however, were "no data", simply reflecting the fact that not all munitions were present in each of the four datasets.

The DEMs generated at each site captured the relative elevations of the munitions themselves, but also local relief of the seabed as well as sessile invertebrates colonizing the seabed (left panels of Figs. 1, 2). Note that some of the muntions were readily apparent in the DEM images but some were not. Where the local topography was flat relative to the height of the target the munitions stand out, but where local relief was large, the targets were not obvious from elevation alone (Figs. 1, 2). For example, the munitions were mofe obvious in the DEM images from the seagrass site (Fig. 2), which had lower relief than the reef site (Fig. 1).

One of the byproducts of generating the DEM for each image sequence during the 2.5-D processing was the creation of an image mosaic from the sequence of overlapping input images (center panels Figs. 1, 2). Note that where the munitions had high color contrast with the background they stood out in the image mosaics. For example, some of the dark targets were obvious against the light sand in both sites. On the other hand, some of the munitions blended in with the color of their surroundings and were not apparent on the mosaics at the scale reproduced here.

Qualitatively, displaying the mosaic as grayscale with pixels classified as munitions in red (right panels Figs. 1, 2) supported the quantitative results reflected in the users and producers accuracy (Tables III-VI). Specifically, the munitions were well discriminated from the background, with a few more false positives apparent in the Reef 1 site than the Seagrass 1 site.

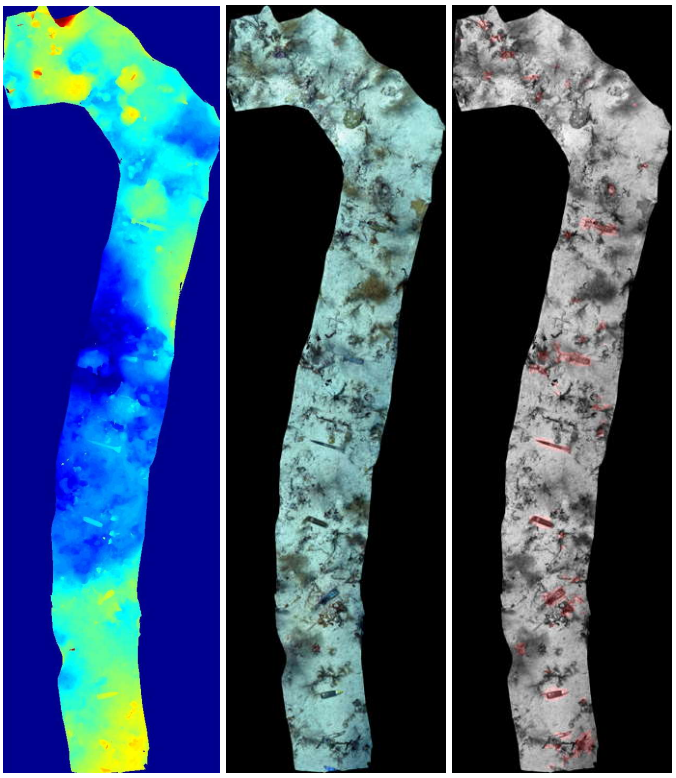

Fig. 1. 2.5-D products for the Reef 1 site. Left: DEM with low (blue) to high (red) colormap. Center: Mosaic of the site. Right: Greayscale of mosaic with pixels classified as munitons colored red.

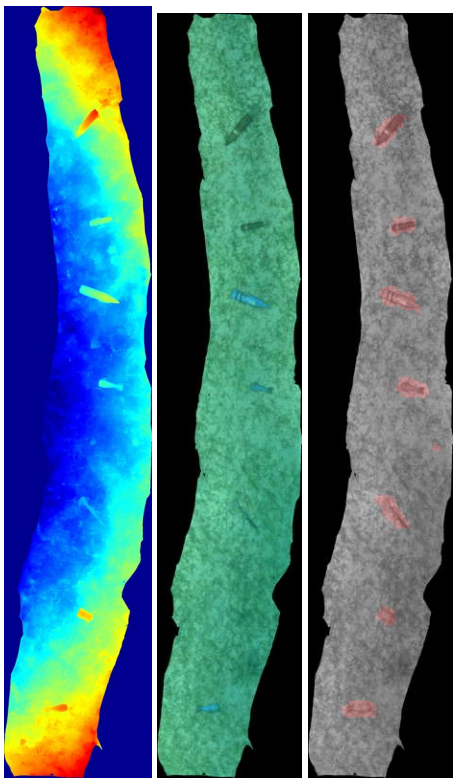

Fig. 2. 2.5-D products for the Seagrass 1 site. Left: DEM with low (blue) to high (red) colormap. Center: Mosaic of the site. Right: Greayscale of mosaic with pixels classified as munitons colored red.

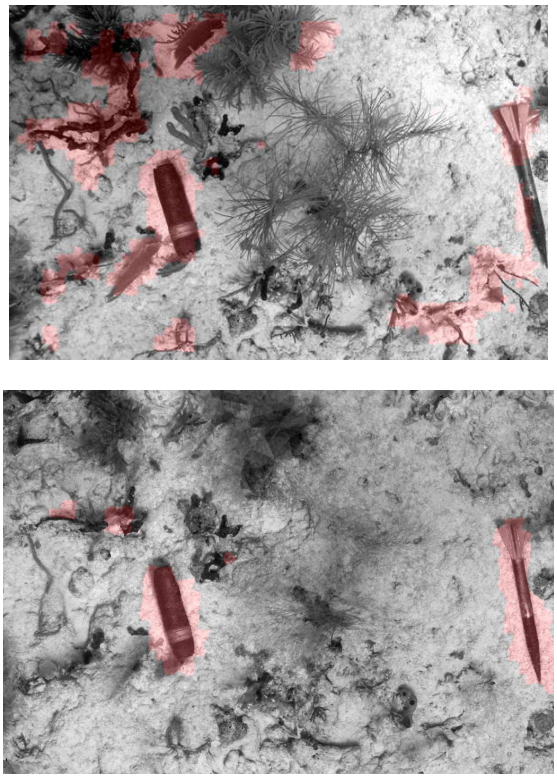

Fig. 3. One of the frames used for the Reef 1 mosaic presented in grayscale. Pixels have been colored red where they were classified as munitions by the 2-D algorithm (top) and the 2.5-D algorithm (bottom). Note fewer false posistives in the 2.5-D results.

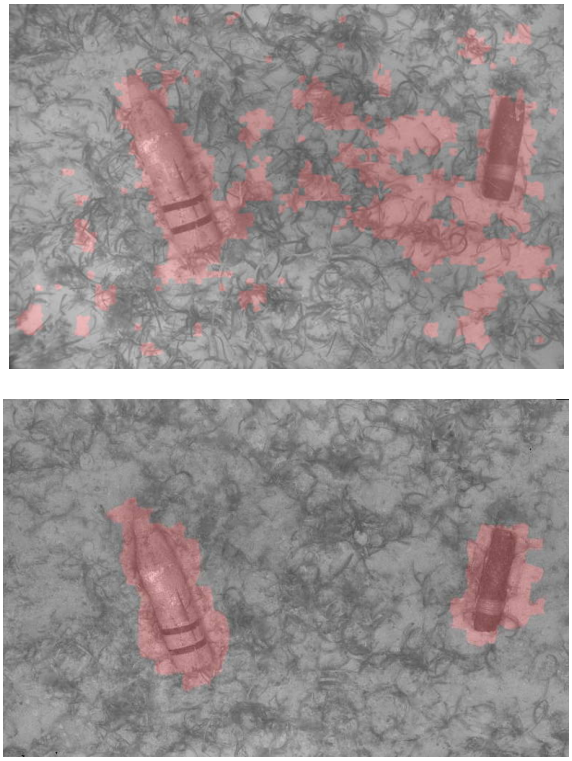

Fig. 4. One of the frames used for the Seagrass 1 mosaic presented in grayscale. Pixels have been colored red where they were classified as munitions by the 2-D algorithm (top) and the 2.5-D algorithm (bottom). Note fewer false posistives in the 2.5-D results.

TABLE V. USERS ACCURACY: DISCRIMINATING MUNITIONS TYPES

| | *Reef 1* | | *Reef 2* | | *Seagrass 1* | | *Seagrass 2* | |
|---|---|---|---|---|---|---|---|---|
| **Class**[b] | ***2-D*** | ***2.5-D*** | ***2-D*** | ***2.5-D*** | ***2-D*** | ***2.5-D*** | ***2-D*** | ***2.5-D*** |
| Seabed | 94 | 98 | 97 | 98 | 98 | 97 | 98 | 98 |
| MA | | | | | | | 0 | 80 |
| MB | | | | | | | 6 | 25 |
| MC | 7 | 63 | 71 | 70 | 49 | 88 | | |
| MD | | | | | | | 0 | 71 |
| ME | 0 | 78 | | | | | | |
| MF | 0 | 93 | 22 | 79 | 2 | 91 | | |
| MG | 0 | 88 | 14 | 69 | 0 | 80 | | |
| MH | 84 | 68 | 0 | 86 | | | 52 | 88 |
| MI | | | | | | | 0 | 67 |
| MJ | | | | | 7 | 71 | | |
| MK | | | 87 | 86 | 29 | 80 | | |
| ML | | | | | 26 | 92 | | |

[b.] Class abbreviations were defined in Table I.

TABLE VI. PRODUCERS ACCURACY: DISCRIMINATING MUNITIONS TYPES

| | *Reef 1* | | *Reef 2* | | *Seagrass 1* | | *Seagrass 2* | |
|---|---|---|---|---|---|---|---|---|
| **Class**[c] | ***2-D*** | ***2.5-D*** | ***2-D*** | ***2.5-D*** | ***2-D*** | ***2.5-D*** | ***2-D*** | ***2.5-D*** |
| Seabed | 83 | 94 | 74 | 97 | 70 | 97 | 57 | 89 |
| MA | | | | | | | 0 | 89 |
| MB | | | | | | | 3 | 95 |
| MC | 45 | 96 | 60 | 92 | 27 | 95 | | |
| MD | | | | | | | 0 | 90 |
| ME | 0 | 88 | | | | | | |
| MF | 0 | 84 | 62 | 74 | 5 | 76 | | |
| MG | 0 | 89 | 64 | 79 | 0 | 84 | | |
| MH | 26 | 92 | 0 | 85 | | | 55 | 73 |
| MI | | | | | | | 2 | 95 |
| MJ | | | | | 17 | 81 | | |
| MK | | | 54 | 83 | 45 | 86 | | |
| ML | | | | | 75 | 86 | | |

[c.] Class abbreviations were defined in Table I.

The qualitative differences between the 2-D and 2.5-D results were apparent when inspecting individual grayscale images with pixels classified as muitions colored in red (Figs. 3, 4). In both the reef (Fig. 3) and seagrass (Fig. 4) cases, there were many more false positive pixels in the 2-D results (top image, Figs. 3, 4) than in the 2.5-D results (bottom images, Figs. 3, 4).

## CONCLUSION

The 2-D image classifier by itself was shown to distinguish munitions from non-munitions (background) with generally high accuracy, but it did not always achieve > 80% accuracy, however, even with a binary munitions/non-munitions scheme. Many false positives were observed at both reef and seagrass sites.

Incorporating additional features based on local relief greatly improved the classification results. Using the 2.5-D information along with the 2-D color and texure features reduced the number of false positives for munitions, increasing overall accuracy from the 50-75% range to 90-95% range (Table II). The reduction in false positives was dramatic, improving users accuracy for muntions from 18-25% to 54-86% (Table III).

Improved accuracy with the 2.5-D data relative to the 2-D data was observed not only on the basic, binary munitions / non-munitions classes. The 2.5-D algorithm also improved discrimination of different types of munitions from one another. The impromentin target identification, as opposed to just presence, was dramatic. For many of the targets accuracy increased from 0% to 70% or greater (Tables V, VI).

Texure and color have been the basis for many recent seabed image classification algorithms [1-5]. As shown here, however, great improvements in accuracy were gained by adding one extra source of information (2.5-D data). These results suggest that adding other additional data, such as object morphology derived from the images themselves and other data derived from complementary sensors, would be a logical next step for further accuracy improvement.

## REFERENCES


[1] N. Pican, E. Trucco, M. Ross, D. M. Lane, Y. Petillot, and I. T. Ruiz, "Texture analysis for seabed classification: co-occurrence matrices vs. self-organizing maps," in *Proceedings of the IEEE Oceans Conference (OCEANS '98)*, 1998, pp. 424-428.

[2] M. S. A. Marcos, L. David, E. Penaflor, V. Ticzon, and M. Soriano, "Automated benthic counting of living and non-living components in Ngedarrak Reef, Palau via subsurface underwater video," *Environmental Monitoring and Assessment,* vol. 145, pp. 177-184, Oct 2008.

[3] O. Pizarro, P. Rigby, M. Johnson-Roberson, S. B. Williams, and J. Colquhoun, "Towards image-based marine habitat classification," in *Proceedings of the IEEE Oceans Conference (OCEANS 2008)*, 2008, pp. 1-7.

[4] M. D. Stokes and G. B. Deane, "Automated processing of coral reef benthic images," *Limnology and Oceanography-Methods,* vol. 7, pp. 157-168, Feb 2009.

[5] O. Beijbom, P. J. Edmunds, D. I. Kline, B. G. Mitchell, and D. Kriegman, "Automated annotation of coral reef survey images," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1170-1177.

[6] A. S. M. Shihavuddin, N. Gracias, R. Garcia, A. Gleason, and B. Gintert, "Image-Based Coral Reef Classification and Thematic Mapping," *Remote Sensing,* vol. 5, pp. 1809-1841, 2013.

[7] A. S. M. Shihavuddin, N. Gracias, R. Garcia, R. Campos, A. C. R. Gleason, and B. Gintert, "Automated Detection of Underwater Military Munitions Using Fusion of 2-D and 2.5-D Features from Optical Imagery," *Marine Technology Society Journal,* vol. 48, pp. 61-71, 2014.

[8] D. Keren and C. Gotsman, "Fitting curves and surfaces with constrained implicit polynomials," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, pp. 31-41, 1999.

[9] G. Hetzel, B. Leibe, P. Levi, and B. Schiele, "3D object recognition from range images using local feature histograms," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. 394-399.

[10] A. Friedman, O. Pizarro, and S. Williams, "Rugosity, slope and aspect from bathymetric stereo image reconstructions," in *Proceedings of the IEEE Oceans Conference, OCEANS 2010*, Sydney, Australia, 2010, pp. 1-9.

[11] Y. Wang, B. S. Peterson, and L. H. Staib, "Shape-based 3-D surface correspondence using geodesics and local geometry," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000, pp. 644-651.

[12] P. Kovesi, "Symmetry and asymmetry from local phase," in *Proceedings of the Tenth Australian Joint Convergence on Artificial Intelligence*, 1997, pp. 2-4.

[13] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton: Lewis Publishers, 1999.