Detecting Underwater Objects

Guillaume Labbe-Morissette

Sylvain Gautier

Labbe-Morissette and Gautier describe a new method to detect underwater regions of interest in real-time side scan sonar imagery.

Who should read this paper?

Sonar operators, hydrographers, biologists, environmentalists, academics, archaeologists, and industrials wishing to automate underwater object detection will be interested in this paper.

Why is it important?

The paper describes an algorithm that can be applied to a large number of detection scenarios that are yet to be explored. It also quantifiably demonstrates promising results for further research, optimization, and comparison by releasing a reference implementation in the public domain in order to foster collaboration and open science.

The technique described allows users to rapidly detect unknown objects and regions of interest from side scan sonar images in order to increase situational awareness, increase the amount of automation, and ultimately reduce costs and delays between data acquisition, interpretation, and actionable results. The paper also shows two major use cases. The first use case is underwater archaeology where automatically finding shipwrecks is demonstrated to be feasible. The second application is in the field of ocean waste management; more specifically, ghost fishing gear, where it shows that detecting lost, abandoned, or derelict fishing gear is also technically feasible. While still in beta, a reference implementation is available as part of the open source side scan sonar toolkit at http://opensidescan.cidco.ca.

About the authors

Guillaume Labbe-Morissette studied mathematics at the University of Montreal and transitioned to software engineering at the University of Quebec in Montreal where he obtained his B.Sc. In 2010, he founded Omnibus Technologies, one of the first data science and distributed computing companies in Canada, where he developed operations research software solutions for the medical, logistics, manufacturing, scientific, and retail sectors. His research interests include artificial intelligence, machine learning, embedded systems, and cybersecurity. He is currently a director at CIDCO – Development Center for Ocean Mapping, a nonprofit research centre dedicated to hydrographic research and development based in Rimouski, QC, Canada. Sylvain Gautier graduated in oceanography from the Institut national des sciences et techniques de la mer (INTECHMER) in Cherbourg (France) and Glamorgan University (United Kingdom) in 2008. He obtained his M.Sc. in underwater acoustics and marine ecology in 2012 at the University of Quebec in Rimouski and has developed expertise in sensor integration, calibration, data acquisition, and processing for hydrographic and geophysical systems. He is currently employed as a marine geomatician at the Canadian Hydrographic Service. His research interests include hydrography, benthic habitat classification, underwater acoustics, and submerged infrastructure inspection.

UNSUPERVISED EXTRACTION OF UNDERWATER REGIONS OF INTEREST IN SIDE SCAN SONAR IMAGERY

Guillaume Labbe-Morissette and Sylvain Gautier

CIDCO – Development Center for Ocean Mapping, Rimouski, QC, Canada

ABSTRACT

This paper details a new method to detect underwater regions of interest (ROI) in real-time side scan sonar imagery, with examples of applications for underwater archaeology and ocean waste management. First, images are synthesized from sonar data, then 2D feature detection algorithms are used to generate point clouds of descriptive visual micro features such as corners and edges. Finally, a clustering algorithm is run on the feature point clouds to detect regions of higher visual feature density. It will be shown that these regions correspond to regions of interest and submerged objects.

KEYWORDS

Hydrography; Artificial intelligence; Computer vision; Pattern recognition; Side scan sonar; Underwater archaeology; Ghost fishing gear

INTRODUCTION

The problem of detecting underwater objects is a recurring issue in many fields, such as hydrography, search and rescue (SAR), underwater archaeology, marine sciences, and many more. Unfortunately, the hostile nature of the underwater environment for human beings, the weak penetration of light, and the difficulty of acquiring high-quality images, along with the high mobilizing costs of scuba or remotely-operated solutions make this endeavour difficult to fulfill.

The shift towards autonomous vehicles equipped with acoustic imaging technology as force multipliers brings along new problems with the multiplication of data sources and produced data. This explosion of data justifies the need for real-time automation to cut costs and delays between data acquisition, interpretation, and actionable results. State of the art autonomous vehicle smart search techniques still require human interaction in the image analysis pipeline to identify areas to investigate [Rutledge et al., 2018].

Furthermore, not all object detection and recognition technologies are suitable for all applications. First, many disciplines rely on information that is inferred through indirect evidence. For example, archaeology is mainly driven by debris, traces, and/or broken vessel parts that would be extremely hard and resource intensive to train a classifier to do. Current trends exploit a basic descriptor and use machine learning, bootstrapping, and other techniques to introduce randomness in matching the training sample so that noisy, imperfect, and partial signatures can be picked up from the noise by finely tuned detectors. Such techniques happen to be successful in scenarios with low geometric variability – for example, the unearthing of Saka tombs in Central Asia – usually discovered under regularly sized round mounds and now found by a convolutional neural network (CNN) fed with satellite imagery [Caspari and Crespo, 2019].

Plenty of work has been done into the area of detecting and recognizing known objects inside images using various descriptors, such as scale-invariant feature transforms (SIFT) [Lowe, 1999], speeded up robust features (SURF) [Bay et al., 2008], binary robust independent elementary features (BRIEF) [Calonder et al., 2010], and even compound methods such as oriented features from accelerated segment test (FAST) and rotated BRIEF (ORB) [Rublee et al., 2011]. Unfortunately, all of these methods require a priori knowledge of the objects to be found and a preliminary training stage using known data in order to adequately detect those objects in new data. More recently, the rise in popularity of CNN methods has given birth to many interesting classifiers and detectors such as AlexNet [Krizhevsky et al., 2012] and Visual Geometry Group (VGG) [Simonyan and Zisserman, 2015] to name only a few who, while sporting remarkable figures, still suffer from the a priori knowledge pitfall and require non-negligible training time and resources.

Finally, many other disciplines rely on finding objects that fall into an unquantifiable amount of categories. For example, ocean waste management has to deal with the detection and retrieval of waste objects that can range from consumer goods, household appliances, vehicles, ghost fishing gear, industrial containers, and many more. As such, a generic waste detector based on object detection and classification would be highly impractical.

The algorithm described in this paper aims to cater to those pitfalls by using an unsupervised approach for domains where direct recognition is not always possible or warranted, either because of the difficulty of accumulating enough data to reliably train CNNs or because of the overwhelming diversity of possible targets. The algorithm builds upon the premises exposed in Viola and Jones [2004] that objects can be represented as cascades of elementary features. It also uses elements found in Fakiris and Papatheodorou [2010], and Masetti and Calder [2012] of using clusters of features found in feature point clouds to detect regions of interest. However, the method disgresses from the use of Haar-like features, local binary features, or grey-level co-occurrence matrix statistics (GLCM) as feature descriptors, and instead uses more robust geometrical features such as FAST [Rosten and Drummond, 2006] and maximally stable external regions (MSER) [Matas et al., 2002; Nistér and Stewénius, 2008]. The clustering algorithm was also changed to density-based spatial clustering of applications with noise (DBSCAN) [Ester et al., 1996] to include noise-rejection, avoid the cluster count problem of k-means, and do away with the need for principal or independent component analysis (PCA/ICA).

Along with benchmarks, two case studies of applications in the field are provided, one where underwater archaeologists need to detect hardly categorizable ship debris and parts, and another one where ocean waste managers need to detect the countless models of fishing gear abandoned or lost at sea. We will show that our method allows users to rapidly detect unknown objects and regions of interest from side scan sonar images in order to increase situational awareness, increase the amount of automation, and ultimately reduce costs and delays between data acquisition, interpretation, and actionable results.

METHODOLOGY

A three-stage workflow was devised to quickly go from raw sensor stream data to actionable object information (Figure 1). In the first phase, images are synthesized from XTF data, which can be either streamed directly through a network connection for real-time analysis or bundled into files by a data acquisition system for post-processing purposes. Either way, the resulting output is a geometrically and radiometrically corrected image suitable for automated analysis.

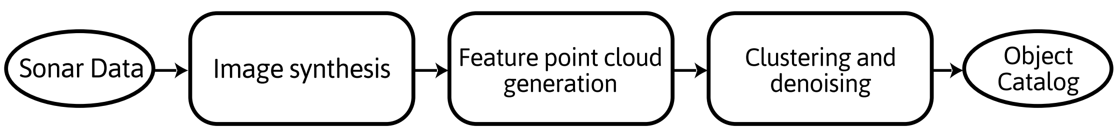

Figure 1: The complete data processing workflow.

In a second phase, visual feature point clouds are generated to detect areas of interest. Since it has been demonstrated that objects in images can be expressed as a large set of smaller visual features [Viola and Jones, 2004], it can be reasonably expected to see dense feature clusters to appear in regions where objects are present, along with a fairly substantial amount of noise.

Consequently, in a third phase, a clustering algorithm is run on the feature point cloud to find areas of higher feature density, which directly correlate with the presence of objects in the image. While the choice of a noise-tolerant clustering algorithm such as DBSCAN allows a single pass, this stage could be broken down into two separate tasks of outlier-rejection and clustering, respectively. Finally, computing the centroid of each feature cluster yields a well-defined and easily georeferenced ROI for each cluster.

Image Synthesis

Due to its fast data acquisition rate, wide area of surveying, relatively low price point, and relative ease of deployment, the side scan sonar has traditionally been a staple for rapidly imaging large bodies of water [Blondel, 2009], especially in fields such as archaeology [Klein, 2002]. The sonar sends acoustic waves through the water and takes signal strength measurements of the received echoes, either amplitude or phase-based, through several transducer arrays known as channels. Each channel receives sequences of vectors of quantized echo samples of the form:

$$\vec{ping} = \{sample_1, ..., sample_n\} \quad (1)$$

The sample count and resolution vary from model to model, but as such they can be used as row pixels by using their value as the pixel intensity, while the vectors can be stacked vertically to generate a full-size greyscale image of the channel. Typically, side scan sonars provide at least two channels, port and starboard.

$$Image_{\mathrm{n \times m}} = \begin{bmatrix} ping_1 \\ ping_2 \\ \dots \\ ping_m \end{bmatrix} = \begin{bmatrix} sample_{1,1} & \dots & sample_{1,\mathrm{n}} \\ sample_{2,1} & \dots & sample_{2,\mathrm{n}} \\ \dots & \dots & \dots \\ sample_{\mathrm{m},1} & \dots & sample_{\mathrm{m,n}} \end{bmatrix} \tag{2}$$

Assuming a stable surveying platform, few corrections are necessary to synthesize intelligible images suitable for automated analysis. To keep preprocessing time to a minimum, we only apply slant-range correction and histogram equalization [Blondel, 2009] (Figure 2).

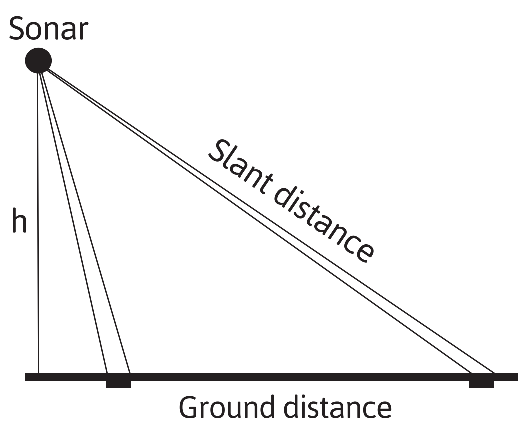

Figure 2: Slant-range scale distortion.

Slant-range Correction

Side scan sonar data contains noticeable visual aberrations due to slant-range scale distortion (Figure 3), which causes identically sized objects to vary in size depending on their distance from the sonar (Figure 2). The correction between the true distance along the ground as a function of the distance along the slant can be found through the following equation [Blondel, 2009].

$$Distance_{\mathrm{Ground}} = \sqrt{Distance^2_{\mathrm{Slant}} - h^2} \tag{3}$$

with h being the height of the sonar taken at the nadir, and the slant distance either obtained directly through XTF data or computed as follows:

$$Distance_{\mathrm{Slant}} = \frac{ct_{\mathrm{twtt}}}{2} \tag{4}$$

using the sound speed as c and t_{twtt} as the two-way travel time of the acoustic beam from the sonar to the bottom. Should h be unavailable, it can be computed using the sonar beam's tilt and roll angles (if available) along with the slant or ground range through elementary trigonometry.

Histogram Equalization

The histogram equalization technique is a method to enhance the contrast in an image [Blondel, 2009]. The process implies mapping the image's intensity histogram to another distribution with a wider and more uniform distribution of intensity values such that the distribution covers the entire range of image values (Figure 4). This transform is readily available in OpenCV [Bradski, 2000].

Feature Cloud Generation

The algorithm rests on the premise that the density of visual features increases inside regions of interest. While most of the Harris [Harris and Stephens, 1988] or Smallest Univalue Segment Assimilating Nucleus (SUSAN) [Smith and Brady, 1995] family of feature detectors could be suitable,

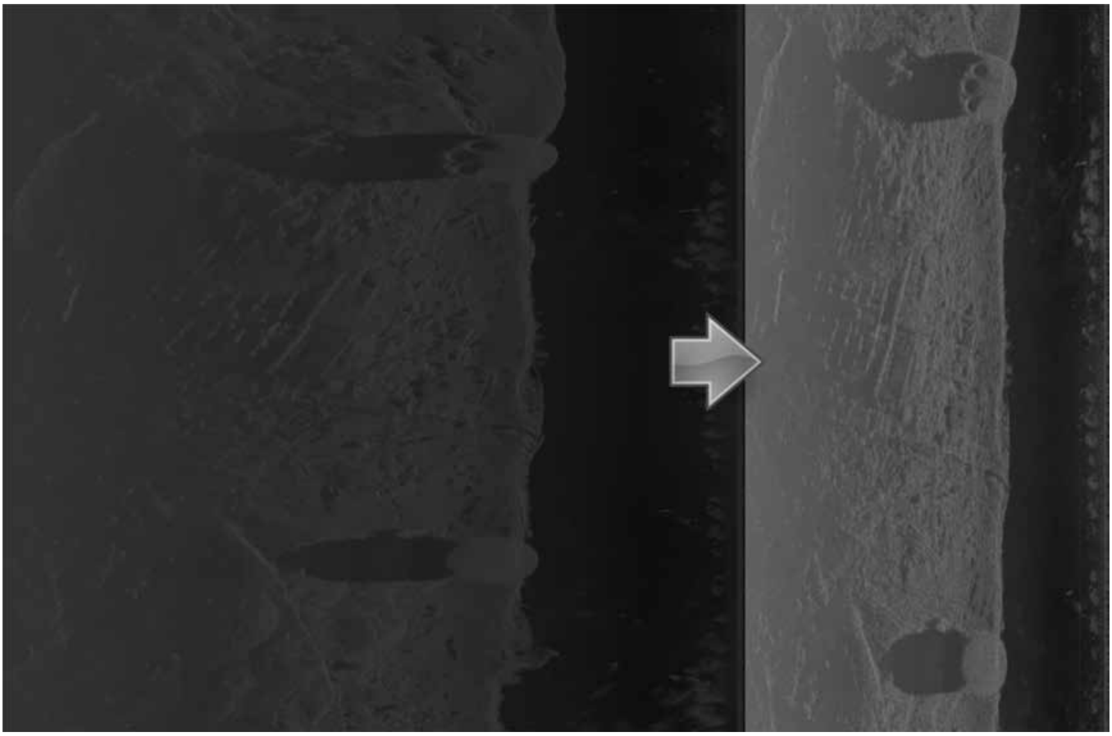

Figure 3: Slant-range scale correction.

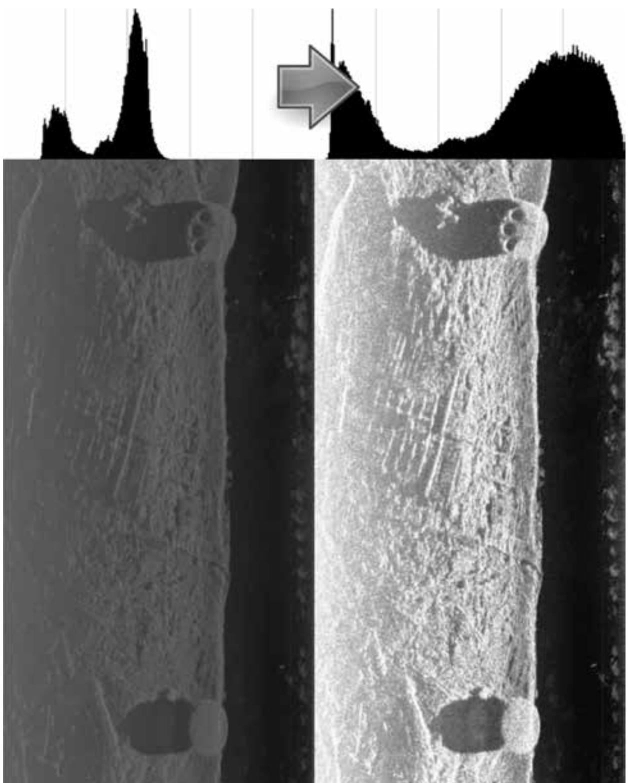

Figure 4: Contrast improvement through histogram equalization clearly enhances the definition and visibility of the shipwreck's boilers.

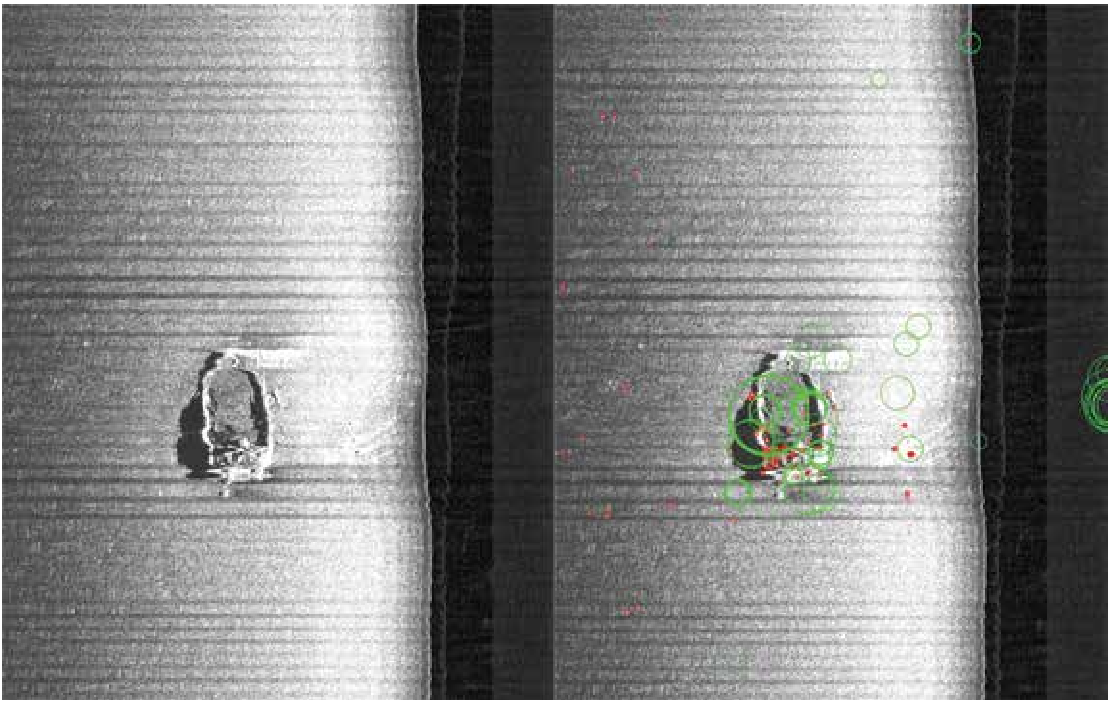

Figure 5: Features from accelerated segment test (FAST) (red) and maximally stable external regions (MSER) (green) features showing strong clusters around the region of interest containing the shipwreck and its nearby debris field.

the FAST [Rosten and Drummond, 2005; 2006] and MSER [Nistér and Stewénius, 2008] algorithms have displayed improved speed of execution, making them especially suitable for embedding into autonomous systems for real-time analysis. Furthermore, since they respectively describe the complementary concepts of corners and colour blobs, it intuitively follows that the combination of both should capture the image's objects' geometry better than a single type of features (Figure 5).

Clustering and Denoising

While many clustering algorithms would be adequate, DBSCAN [Ester et al., 1996] provides a quick and practical solution to our clustering need due to its ease of use, its ability to reject noise-like features, and its support for an arbitrary number of clusters without a priori knowledge. This allows for rapidly searching the sonar images for an arbitrary amount of clusters while discarding noise at the same time (Figure 6). Once the point cloud is clustered, the centroid can be used to define a bounding box for the ROI.

PERFORMANCE

Calibration Parameters

The algorithm is sensitive to several parameters that can be tuned to optimize the performance for a given application. Namely, feature generation parameters (FAST and MSER) allow finely-grained control over what kind of features should be noticed. Furthermore, the clustering parameters (DBSCAN) allow flexible aggregation of the detection results based on the density of the features and their surroundings.

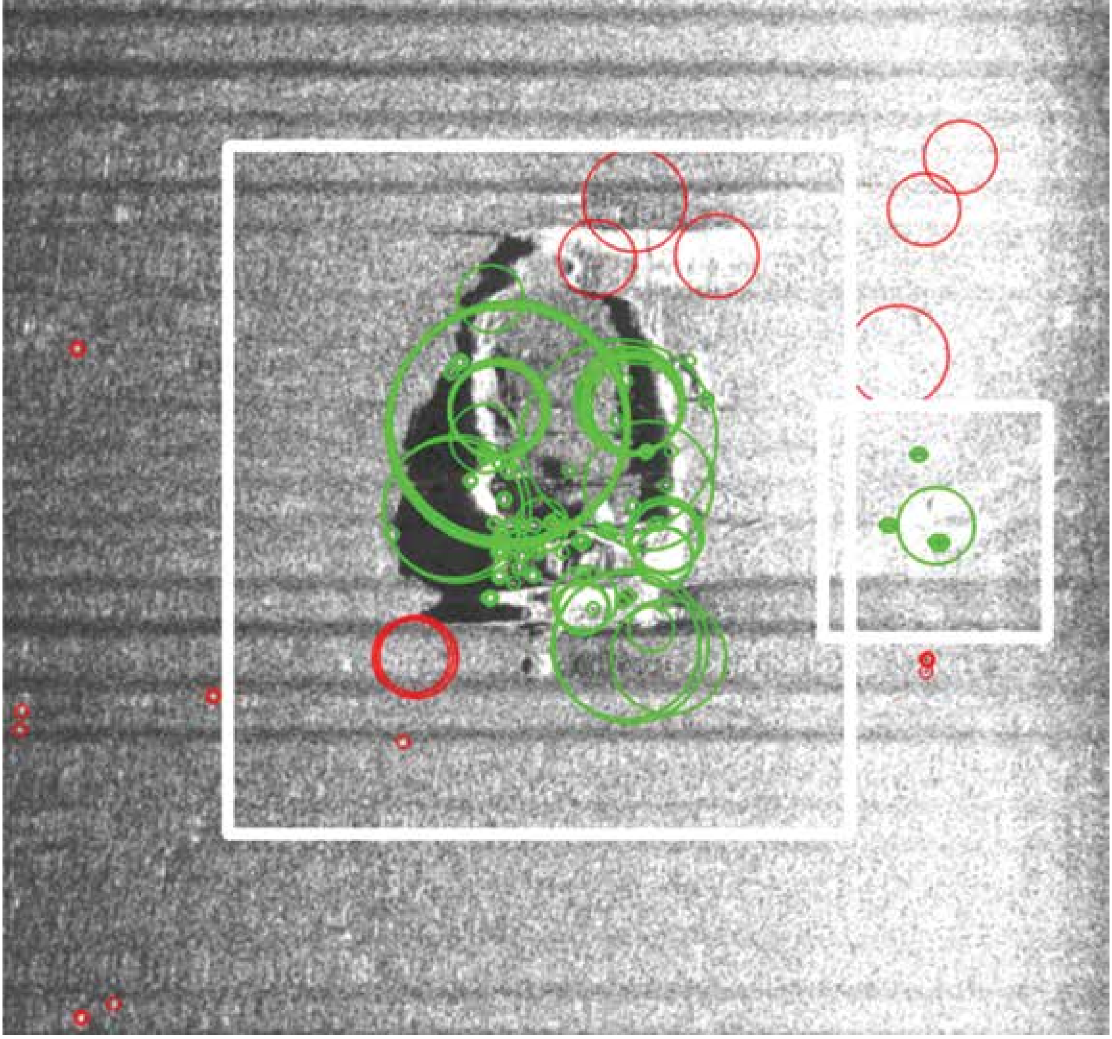

Figure 6: Effective noise rejection through feature clustering, with region of interest (ROI) bounding boxes (white), clustered features (green), and features rejected as noise (red).

FAST Threshold

In the FAST feature generation algorithm, it defines the intensity threshold value t such that a given pixel p is a corner if there exists a set of 12 contiguous pixels on the 16 pixel Bresenham circle around p, whose intensity is all larger (brighter) than I(p) + t or all lower (darker) than I(p) - t [Rosten and Drummond, 2006].

MSER Delta

In the MSER feature generation algorithm, the delta parameter is used when determining if a region of connected components is maximally stable [Matas et al., 2002; Nistér and Stewénius, 2008].

MSER Minimum Area

In the MSER feature generation algorithm, the minimum area parameter is the minimum pixel count for a region to be considered a MSER [Matas et al., 2002; Nistér and Stewénius, 2008].

DBSCAN Epsilon

In the DBSCAN clustering algorithm, the epsilon parameter is the maximum distance between neighbouring features [Ester et al., 1996].

DBSCAN Minimum Points

In the DBSCAN clustering algorithm, the minimum points parameter is the minimum number of neighbouring features for a feature

Dataset	Precision (%)	Recall (%)
Shipwrecks	100	100
Crab Traps	72.7	100

Table 1: Dataset benchmarks.

to be considered part of a cluster (zone of interest) [Ester et al., 1996].

Benchmarks

The benchmarking methodology involved two datasets, one consisting of plane and ship wrecks (n=10) and another one with lost/abandoned fishing gear (n=6). Both have been manually investigated by a human operator who identified regions of interest and pinpointed their centres. Comparison with the automatic detection results yielded the precision and recall figures contained in Table 1. A genetic algorithm was created to fit optimal parameters on the two datasets, using the sum of precision and recall as a fitness function.

One of the main takeaways from Table 1 is the interestingly high recall figures. This shows a clear aptitude at sifting through large quantities of data and highlighting regions of interest without missing any detections while keeping the false-positive rate relatively low, if any.

APPLICATIONS

Automatically detecting and mapping underwater regions of interest yields many interesting applications in terms of surveying automation and generating added value from sonar data.

Underwater Archaeology

In August 2019, IRHMAS and CIDCO were involved in an archaeological campaign to find remains of the SS *Germanicus*, a large cargo steamer that ran aground in 1919, and of the SS *Scotsman*, one of the oldest shipwrecks in the Saint Lawrence seaway, sunk in 1846, both located near Le Bic, Canada (Figure 7).

The presence of large salient objects such as large hull pieces, boilers, and large debris fields made this a perfect training ground for the automatic detection algorithm, which detected the large components without issues, even against irregular backgrounds with numerous features (Figure 8).

Ocean Waste Management

In 2018, CIDCO, based on advisories from Fisheries and Oceans Canada, started testing different methodologies for detecting ghost fishing gear to facilitate retrieval operations.

Ghost fishing gear is defined as fishing equipment – either abandoned, lost, or thrown away – that continues to be functional in the water, therefore continuing to fulfill its function of trapping, mutilating, or killing marine life in an unsupervised fashion. Encounters and entanglement with ghost fishing gear are among the main lethal threats for many endangered species, such as large mammals who can easily become entangled in vertical ropes suspended in the water. For example, more than 80% of right whales in the North Atlantic (*Eubalaena glacialis*) become

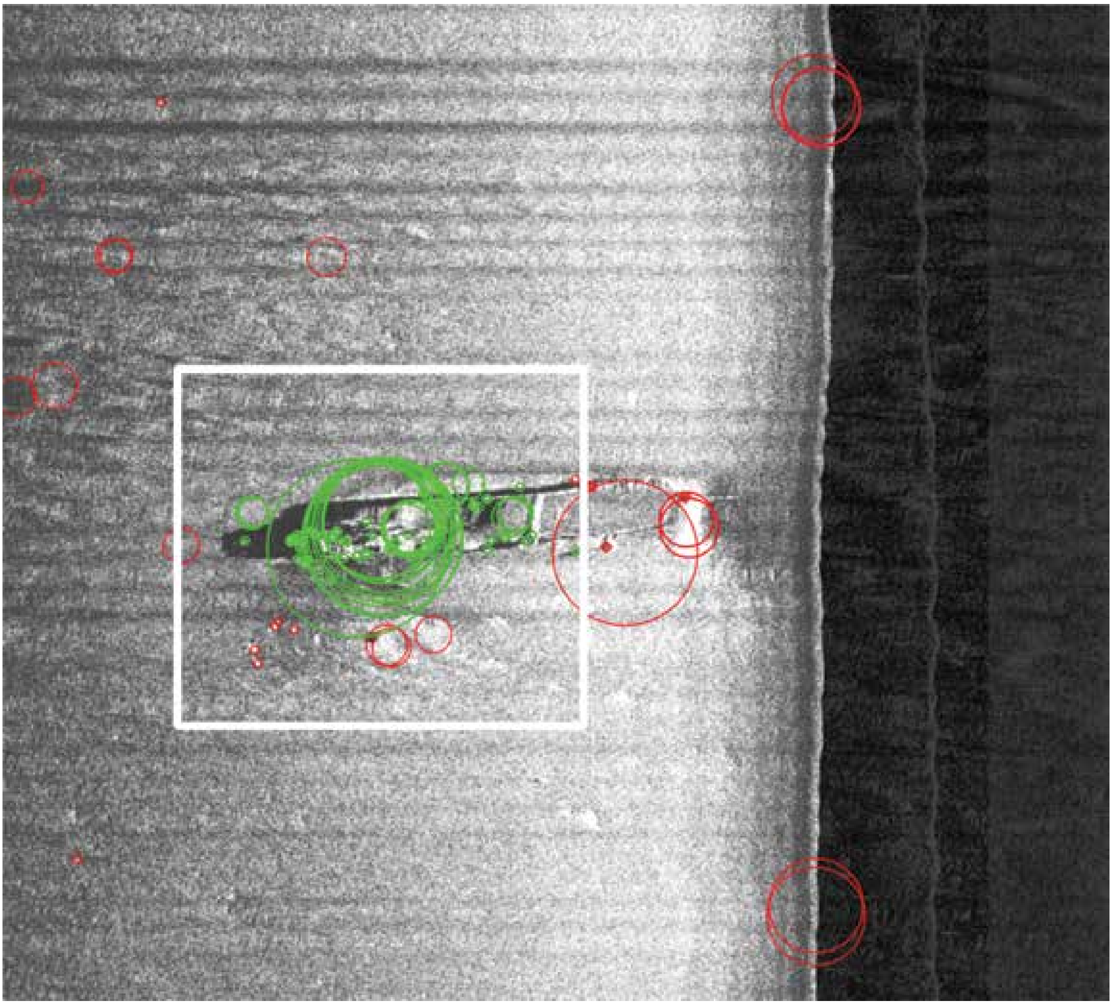

Figure 7: The wreck of the SS *Scotsman*.

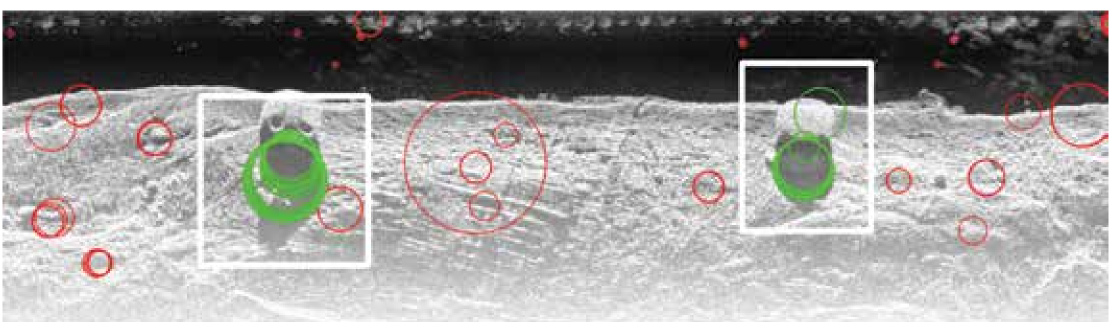

Figure 8: Boilers from the SS *Germanicus* surveyed with a StarFish 990F side scan sonar.

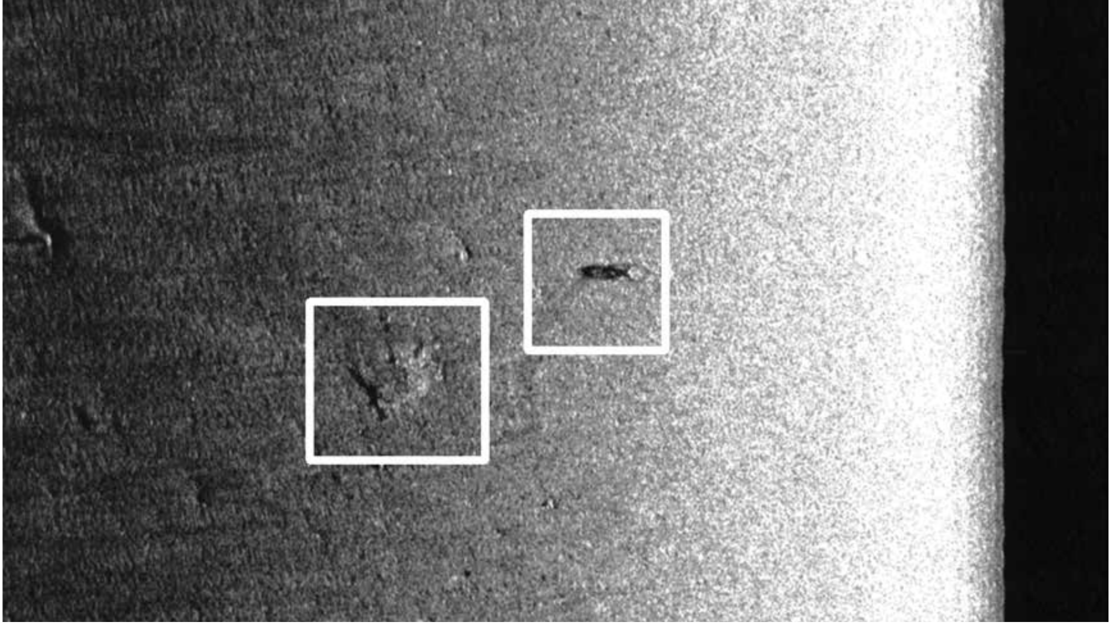

Figure 9: A crab trap along with its rope hanging in the water column scanned using an Edgetech 272-TD analog side scan sonar.

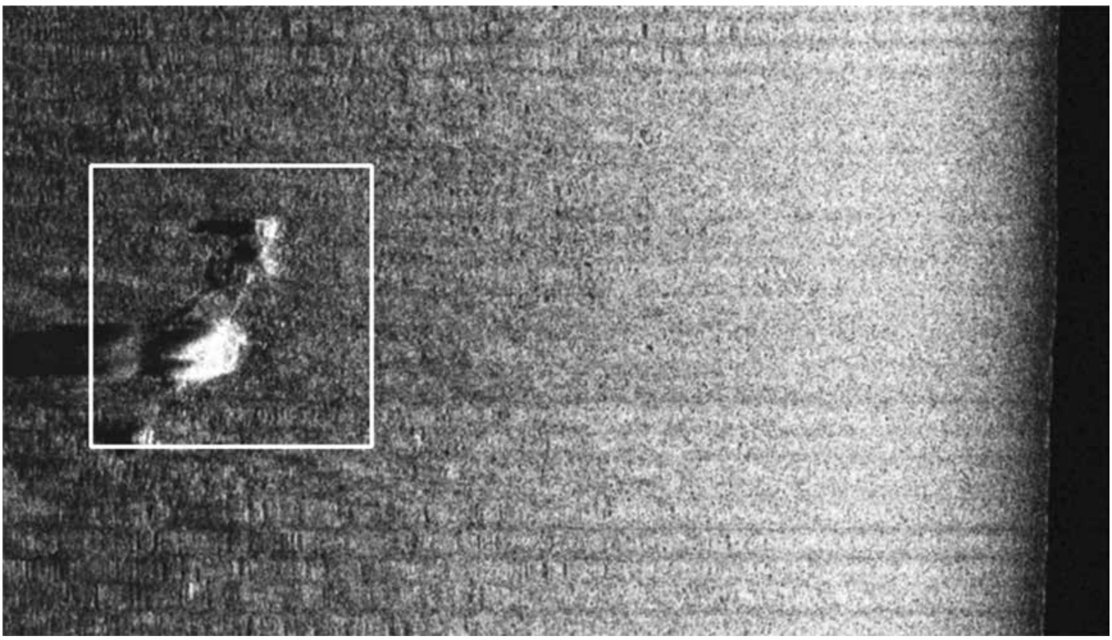

Figure 10: Abandoned aquaculture long-lines with sunken concrete blocks scanned with a Klein 3000 side scan sonar.

entangled in fishing gear at least once in their lifetimes [Knowlton et al., 2012].

The highly variable geometry of fishing gear such as crab pot models (Figure 9), nets, or long-lines (Figure 10) makes a strong case for the relevance of unsupervised algorithms such as this one in the context of ghost fishing gear detection.

DISCUSSION

The algorithm definitely shows promising performance, especially with its high recall figures. However, this could be influenced by the small sample size of the test dataset. Additional data could be used to strengthen this claim.

Furthermore, the high recall figures come at the cost of an increased false-positive rate, which, depending on the application, can result in additional costs. Therefore, care should be taken when implementing the algorithm into a process so that false-positives can be efficiently handled. For example, such a process could have a human operator manually weed out false-positives, which would still result in a net gain over manually processing the entire data stream.

Moreover, while side scan technology has been our main focus in the context of this research, the algorithm can be readily generalized to other imaging technologies such as synthetic aperture sonar, multibeam backscatter, satellite imagery, and other imaging technologies by simply adapting the image synthesis phase. This opens interesting possibilities in terms of future applications.

Additionally, an interesting consequence of the high recall figures is a potential use case as an attention mechanism that could be leveraged to focus more specific detectors onto restricted regions of interest, a feature which could have some interesting benefits in scenarios with a lot of noise. For example, it could be used to select regions in a first pass for specialized classifiers such as convolutional neural networks to classify their contents in a second pass. More research in that direction is currently underway.

Finally, while special care has been taken to make the technology embeddable into real-time computers, this feature is untested at the time of writing. However, the use of XTF files, and more specifically, conformity to the XTF file format specification ensures that the data can be readily picked up from a live acquisition system's output files and treated on the fly on board a boat or autonomous vessel exactly the same way that it can be post-processed in the lab from data brought back from the boat at the end of the day.

The goal of this paper is thus to describe the method and quantifiably demonstrate promising results to open the door for public use and further research, optimization, and comparison by releasing a reference implementation in the public domain.

CONCLUSION

The ability to automatically extract underwater regions of interest in large quantities of data brings about new possibilities in terms of underwater surveying automation.

Our algorithm opens up interesting applications in the field of underwater archaeology, such as fully automated underwater inventories that can significantly lower the cost of surveying large areas with high-speed vessels. This allows archaeologists to do more work, and partially frees them from the dependency on hydrographers to post-process their field data.

In the case of ghost fishing gear, field testing of our algorithm has shown efficient detection capabilities. This opens the way for major improvements in the efficiency of ocean cleaning methods by allowing real-time detection and geolocation of waste retrieval zones over large surfaces, thus lowering the costs and data post-processing times compared to a human operator.

ACKNOWLEDGMENTS

The authors would like to thank MERINOV and the IRHMAS for providing CIDCO with the opportunity to acquire the shipwreck and ghost fishing gear data required to conduct the field testing of the algorithm described in this paper. CIDCO would also like to extend its gratitude to Fisheries and Oceans Canada and Canadian Hydrographic Service for their continuous support of hydrographic research and development, along with Technopole Maritime du Quebec for its continuous efforts at bringing ocean research and industry together to help solve tomorrow's marine challenges.

SOFTWARE

A reference implementation under MIT license is available as part of the OpenSidescan toolsuite at http://opensidescan.cidco.ca.

REFERENCES

Bay, H.; Ess, A.; Tuytelaars, T.; and Van Gool, L. [2008]. *Speeded-up robust features (SURF)*. Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346-359.

Blondel, P. [2009]. *The handbook of sidescan sonar*. Praxis.

Bradski, G. [2000]. *The opencv library*. Dr. Dobb's Journal of Software Tools.

Calonder, M.; Lepetit, V.; Strecha, C.; and Fua, P. [2010]. *Brief: binary robust independent elementary features*. In: Proceedings of the 11th European Conference on Computer Vision: Part IV. ECCV'10. Springer-Verlag, Berlin, Heidelberg, pp. 778-792.

Caspari, G. and Crespo, P. [2019]. *Convolutional neural networks for archaeological site detection - finding "princely" tombs*. Journal of Archaeological Science, Vol. 110.

Ester, M.; Kriegel, H.-P.; Sander, J.; and Xu, X. [1996]. *A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise*. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226-231.

Fakiris, E. and Papatheodorou, G. [2010]. *Automatic target detection in side-scan sonar imagery, using feature based acoustic classification and independent component analysis*. In: 4th International Conference and Exhibition on Underwater Acoustic Measurements.

Harris, C. and Stephens, M. [1988]. *A combined corner and edge detector*. In: Proceedings of the Fourth Alvey Vision Conference, pp. 147-151.

Klein, M. [2002]. *International handbook of underwater archaeology*. The Springer Series in Underwater Archaeology.

Knowlton, A.; Hamilton, P.; Marx, M.; Pettis, H.; and Kraus, S. [2012]. *Monitoring North Atlantic right whale eubalaena glacialis entanglement rates: a 30 year retrospective*. Marine Ecology Progress Series, Vol. 466, pp. 293-302.

Krizhevsky, A.; Sutskever, I.; Hinton, G.E. [2012]. *Imagenet classification with deep convolutional neural networks*. In: Proceedings of the 25th International Conference on Neural Information Processing Systems – Volume 1. NIPS'12. Curran Associates Inc., USA, pp. 1097-1105.

Lowe, D.G. [1999]. *Object recognition from local scale-invariant features*. In: Proceedings of the 7th IEEE International Conference on Computer Vision, Vol. 2, pp. 1150-1157.

Masetti, G. and Calder, B. [2012]. *Remote identification of a shipwreck site from MBES backscatter*. Journal of Environmental Management, Vol. 111, pp. 44-52.

Matas, J.; Chum, O.; Urban, M.; and Pajdla, T. [2002]. *Robust wide baseline stereo from maximally stable extremal regions*. In: Rosin, P.L. and Marshall, A.D. (eds.), British Machine Vision Association.

Nistér, D. and Stewénius, H. [2008]. *Linear time maximally stable extremal regions*. In: Proceedings of the 10th European Conference on Computer Vision: Part II. ECCV'08. Springer-Verlag, Berlin, Heidelberg, pp. 183-196.

Rosten, E. and Drummond, T. [2005]. *Fusing points and lines for high performance tracking*. In: IEEE International Conference on Computer Vision, Vol. 2, pp. 1508-1511.

Rosten, E. and Drummond, T. [2006]. *Machine learning for high-speed corner detection*. In: Leonardis, A.; Bischof, H.; and Pinz, A. (eds.), Computer Vision – ECCV'06. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 430-443.

Rublee, E.; Rabaud, V.; Konolige, K.; and Bradski, G. [2011]. *Orb: an efficient alternative to sift or surf*. In: Proceedings of the 2011 International Conference on Computer Vision. ICCV'11. IEEE Computer Society, Washington, DC, USA, pp. 2564-2571.

Rutledge, J.; Yuan, W.; Wu, J.; Freed, S.; Lewis, A.; Wood, Z.; Gambin, T.; and Clark, C. [2018]. *Intelligent shipwreck search using autonomous underwater vehicles*. In: Proceedings of the IEEE International Conference on Robotics and Automation. ICRA 2018.

Simonyan, K. and Zisserman, A. [2015]. *Very deep convolutional networks for large-scale image recognition*. In: International Conference on Learning Representations.

Smith, S.M. and Brady, J.M. [1995]. *SUSAN - a new approach to low level image processing*. International Journal of Computer Vision, Vol. 23, pp. 45-78.

Viola, P. and Jones, M.J. [2004]. *Robust real-time face detection*. International Journal of Computer Vision, Vol. 57, No. 2, pp. 137-154.