

ANOMALY AND TARGET DETECTION IN SYNTHETIC APERTURE SONAR

By

PRINCESS LYONS

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2019

© 2019 Princess Lyons

To Mom and Dad, for your unwavering love, support and faith

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Alina Zare, who inspired me to reach higher and challenge my potential. I would like to also thank my committee members, Dr. Damon Woodard and Dr. Paul Gader. Additionally, the Office of Naval Research (ONR) grant N00014-16-1-2323 who funded this work. I cannot forget to acknowledge my family for their love and prayers and the Lord for the strength to complete this program. Only with your faith and support in my goals, would any of this have been possible. I would like to acknowledge my stress and without you, my dear anxiety, this journey would have been different. To my short attention span, thank you for the hours spent googling random things before starting research just to warm up my mental. A very special thanks to my dear friends Jahnelle Jordan and Joshua K. Peeples for encouraging and accompanying me on this academic endeavor as well as providing excellent advice and entertaining banter to survive.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	4
LIST OF TABLES	7
LIST OF FIGURES	8
ABSTRACT	10
CHAPTER	
1 INTRODUCTION	12
1.1 Anomaly and Target Detection	12
1.2 Problem Statement	14
1.3 Overview of Research	15
2 LITERATURE REVIEW	17
2.1 Overview of SONAR and Synthetic Aperture SONAR (SAS) Imagery	17
2.2 Motivation for Underwater Object Detection	19
2.3 Detection Limitations	20
2.4 Statistics- and Image Processing-Based Methods	22
3 METHOD	25
3.1 Estimate Templates	25
3.2 Detectors: Produce Confidence Maps	27
3.2.1 Matched Filters	29
3.2.2 Adaptive Cosine Estimators	29
3.2.3 Convolution	30
3.2.4 Normalized Cross-Correlation	30
3.2.5 Reed-Xiaoli Detector	31
3.3 Combine Confidence Maps	32
3.4 Non-Maximum Suppression and Detection	33
4 EXPERIMENTS AND DISCUSSION	35
4.1 N-Fold Cross-Validation	35
4.2 Site A*: K-Means vs K-Medoids	36
4.2.1 Original Templates	39
4.2.2 Pruned Templates	45
4.2.3 Summary	50
4.3 Site A*: Combined Frequency	52
4.3.1 Four Original Templates	54
4.3.2 Six Original Templates	58
4.3.3 Summary	61

4.4	All Five Sites	65
4.4.1	Original Templates	66
4.4.2	Pruned Templates	69
4.4.3	Combined Frequency	72
5	CONCLUSION	76
	REFERENCES	82
	BIOGRAPHICAL SKETCH	85

LIST OF TABLES

<u>Table</u>		<u>page</u>
4-1 Template-Based Experimental Setup Parameter Options		35
4-2 Number of images and targets in each fold of Site A*		37
4-3 K-Means 4 Original Templates table: FAR vs TPR		40
4-4 K-Medoids 4 Original Templates table: FAR vs TPR		41
4-5 K-Means 6 Original Templates table: FAR vs TPR		43
4-6 K-Medoids 6 Original Templates table: FAR vs TPR		44
4-7 K-Means 4 Pruned Templates table: FAR vs TPR		46
4-8 K-Medoids 4 Pruned Templates table: FAR vs TPR		47
4-9 K-Means 6 Pruned Templates table: FAR vs TPR		48
4-10 K-Medoids 6 Pruned Templates table: FAR vs TPR		50
4-11 Combined Frequency K-Means 4 Original Templates		63
4-12 Combined Frequency K-Medoids 4 Original Templates		63
4-13 Combined Frequency K-Means 6 Original Templates		64
4-14 Combined Frequency K-Medoids 6 Original Templates		64
4-15 Number of images and targets by site		65
4-16 All Five Sites: High Frequency K-Means 8 Original Templates		67
4-17 All Five Sites: Low Frequency K-Means 8 Original Templates		68
4-18 All Five Sites: High Frequency K-Means 8 Pruned Templates		70
4-19 All Five Sites: Low Frequency K-Means 8 Pruned Templates		71
4-20 All Five Sites: Combined Frequency K-Means 8 Original Templates using Addition		73
4-21 All Five Sites: Combined Frequency K-Means 8 Original Templates using Multiplication		74

LIST OF FIGURES

<u>Figure</u>		<u>page</u>
3-1 Dimensions of a mugshot, target template and detection window		26
3-2 Example of template pruning process		27
3-3 Diagram of how confidence maps are produced		28
3-4 Kernel Dimensions for RX Methods [1]		31
3-5 Example of combining confidence maps		32
3-6 Example of Non-Maximum Suppression		33
4-1 AUV path and fold designation		36
4-2 Example of an Experiment		38
4-3 K-Means 4 Original Templates ROC Curves		40
4-4 K-Medoids 4 Original Templates ROC Curves		41
4-5 K-Means 6 Original Templates ROC Curves		42
4-6 K-Medoids 6 Original Templates ROC Curves		43
4-7 K-Means 4 Pruned Templates ROC Curves		46
4-8 K-Medoids 4 Pruned Templates ROC Curves		47
4-9 K-Means 6 Pruned Templates ROC Curves		48
4-10 K-Medoids 6 Pruned Templates ROC Curves		49
4-11 Combined Confidence maps and Combined Frequency Confidence Maps		53
4-12 K-Means Addition Combined Frequency 4 Original Templates ROC Curves		54
4-13 K-Means Multiplication Combined Frequency 4 Original Templates ROC Curves		55
4-14 K-Medoids Addition Combined Frequency 4 Original Templates ROC Curves		56
4-15 K-Medoids Multiplication Combined Frequency 4 Original Templates ROC Curves		57
4-16 K-Means Addition Combined Frequency 6 Original Templates ROC Curves		58
4-17 K-Means Multiplication Combined Frequency 6 Original Templates ROC Curves		59
4-18 K-Medoids Addition Combined Frequency 6 Original Templates ROC Curves		60
4-19 K-Medoids Multiplication Combined Frequency 6 Original Templates ROC Curves		60

4-20	High Frequency K -Means 8 Original Templates ROC Curves	67
4-21	Low Frequency K -Means 8 Original Templates ROC Curves	68
4-22	High Frequency K -Means 8 Pruned Templates ROC Curves	70
4-23	Low Frequency K -Means 8 Pruned Templates ROC Curves	71
4-24	K -Means Addition Combined Frequency 8 Original Templates ROC Curves	73
4-25	K -Means Multiplication Combined Frequency 8 Original Templates ROC Curves	74

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

ANOMALY AND TARGET DETECTION IN SYNTHETIC APERTURE SONAR

By

Princess Lyons

December 2019

Chair: Alina Zare

Major: Electrical and Computer Engineering

Automated anomaly and target detection are commonly used as a prescreening step within a larger target detection and target classification framework to find regions of interest for further analysis. Many anomaly and target detection algorithms in the literature have been developed for application to target detection in Synthetic Aperture Sonar (SAS) imagery which produces high quality intensity images of underwater scenes. In this paper, a comparison of unsupervised and supervised anomaly and target detection algorithms for target detection in synthetic aperture sonar is presented.

Three different experiments are done on two subsets of a large SAS dataset containing images captured in high frequency and low frequency. Template matching and statistical methods are used in order to detect targets within the SAS imagery. For detection, it is desired that a low number of false alarms be detected with a high number of true positives. This means that a detector is able to distinguish a target from an anomaly well. In practice, it is likely that human intervention is required to confirm targets. If a detector returns a low number of false alarms the amount of labor necessary by an operator is minimal.

The purpose of the first two experiments is to observe the effects of varying certain parameters for detection. Among these parameters are the number of templates, dimensions and frequency of the SAS images and methods used to process the images. For the template matching methods, *K*-Means and *K*-Medoids are employed to generate templates of the targets to be detected. Six different detectors are utilized: adaptive cosine estimator, squared

adaptive cosine estimator, spectral matched filter, matched filter, normalized cross-correlation, and convolution. Two statistical detectors are utilized: Reed-Xiaoli and Reed-Xiaoli Left-Right. The results of this thesis show that template-based detection produces comparable results to statistics-based detectors. The spectral matched filter and Reed-Xiaoli detectors show the most promising results for anomaly and target detection when low frequency data is used.

CHAPTER 1

INTRODUCTION

Distinguishing between anomaly and target detection methods may seem like a simple task. However, they employ very different methodologies. The unsupervised versus supervised nature of anomaly and target detection is studied and exemplified with real life applications in this thesis.

1.1 Anomaly and Target Detection

Anomaly detection is a common prescreening step in target detection applications. The goal of anomaly detection is to locate anything that looks distinct from its surroundings or, in other words, anything that can be categorized as “out of the ordinary” [2]. Things of this nature are considered anomalies, outliers or objects of interest. Anomaly detection methods are employed in various domains including finance, security, medicine, video surveillance, networking, communications, process quality control, hyperspectral imaging and more [3]. These methods usually operate under circumstances where no prior information is available about the objects of interest, deeming it an unsupervised problem. Usually, prior information about the problem domain that aids in distinguishing the objects of interest, also known as background information, is available and incorporated into the detection process.

While anomaly detection is a broad term in general with many developed techniques in the literature, when used for a domain-specific problem the application becomes more tuned and focused. For example, fraud detection is described as an application of anomaly detection where financial activities are monitored for anomalous patterns [4]. In the medical industry, anomaly detection methods have been applied to detect unusual patient-management decisions using a patient’s past electronic health records to generate an alert if such a decision is encountered [5]. An approach that detects abnormal flights from routine airline operations using flight data recorder (FDR) data to determine operational implications was also developed in [6].

Due to the structure of the data in these problems, they can be considered signal processing tasks where the signal is defined as a numerical representation of physical events. Therefore, the same principles used in detection for the aforementioned examples can be applied to pixel intensities in image-based anomaly detection. Kim [7] models samples of network packet header data as frames or images to perform image-based anomaly detection to prevent large-scale network attacks. In hardware security applications, Trojans in electrical components inserted by an untrusted foundry can be detected by analyzing images of the integrated circuits (ICs) [8]. In medical imaging, anomaly detection has been used for years to alert doctors of abnormalities in X-Rays, Computerized Axial Tomography (CAT) scans and Magnetic Resonance Imaging (MRI).

Again, this type of detection is typically done in an unsupervised manner with some background information about the problem but no prior knowledge of the data. If assumptions are made to model these unexpected attacks or Trojans, supervision is used to distinguish these objects of interests as normal or abnormal. This naturally leads one to call this type of detection target detection. In these instances, information is known about the object of interest to be detected and can simply be called a target.

Known information about a detection problem determines the overall difficulty. Problems with no information are inherently more difficult than those with information. This is the key distinction between anomaly and target detection. Accuracy measures often emphasize these differences. Since anomaly detection typically lacks prior knowledge like descriptions of the underwater environment or types of objects or targets, things that aren't actually anomalous may be detected producing higher false alarm rates. There are no specific characterizations of the objects to be detected so all data is considered in the detection which generally results in computationally expensive processing with large datasets. Target detection however uses the prior knowledge to simplify the problem making it supervised. Targets are modeled and these models are used for comparison with the data to perform detection. Both parametric and non-parametric models have been explored in the literature. The presence of prior knowledge

often leads to lower false alarm rates although there is potential for higher false negative rates. However, this added sense of direction in detection can increase processing speeds and confidence of detected targets. Though added information is usually advantageous, there may still be situations where everything about the target is known but not useful. In these situations, prior knowledge can be ineffective and useless if there are no distinguishing or distinct qualities about the target improve detection.

1.2 Problem Statement

Sound navigation and ranging, commonly referred to as SONAR, is a widely used technology for navigating and creating views of underwater scenes. Direct observations of underwater environments and the seafloor are not feasible due to poor visibility. Propagating sound waves however, can communicate with an underwater environment by returning echoes that bounce off of the environment and objects within. Innovation has led to the development of synthetic aperture SONAR (SAS) which is especially useful for underwater object detection as it produces high quality images of underwater scenes. It is an enhancement of standard SONAR technology that involves complex post-processing techniques which result in a higher resolution intensity image of an underwater scene. Underwater mine countermeasures (MCM) make use of these high resolution images to detect and identify dangerous objects. This area of study focuses on safely and effectively locating potentially dangerous objects laying on or buried beneath the seafloor [9]. Often placed with malicious intent, objects such as mines can cause serious damage to nearby submarines, military or civilian ships and even marine life if they are detonated.

When performing detection on high resolution SAS imagery, quick, efficient and accurate methods are a necessity. Variability in the objects of interests, underwater scene environments and large data sizes introduce many difficulties in the detection process. For this reason, supervised and unsupervised methods have been utilized to model variability and flag objects that are out of the ordinary. Nevertheless, deciding which approach is preferable over the other is often done in a trial and error fashion often failing to address the aforementioned needs. In

this thesis, a comparison of unsupervised methods, including the widely popular Reed-Xiaoli (RX) detector [10], and template-based supervised methods is studied using a large dataset comprised of high and low frequency SAS images from five different underwater locations. Various objects and targets are scattered throughout the images although every image does not have a target. Emphasis is placed on highlighting important underlying characteristics of successful supervised and unsupervised anomaly and target detection in SAS imagery containing high varying objects of interests and underwater scene environments.

1.3 Overview of Research

In this thesis, a comparison of methods for efficient and effective detection of anomalies and targets in SAS imagery is conducted. Unsupervised approaches requiring no prior knowledge of the underwater environment or targets and supervised approaches requiring this information, were investigated for this study. Knowledge of the seafloor environments and objects of interest are incorporated into the detection process. The popular and successful unsupervised RX detector was used for comparisons against various supervised template-based detector methods. Statistical and similarity metrics are used to determine the presence of anomalies and targets.

Various algorithms which employ image processing and machine learning techniques are created for experimentation. Focus is placed on developing methods that produce high true positive rates with low false alarm rates for improved accuracy. The goal of this study is to determine the key factors of anomaly and target detection in SAS imagery when using unsupervised versus supervised methods. In many instances, anomaly detection has the tendency to not only find objects of interest but also objects that are not of interest resulting in high false positive rates. Whereas, target detection often fails to detect all objects of interest due to its inability to detect targets unlike *known* objects of interest.

A large dataset of over 6,000 images which contain multiple target types is used. The targets are scattered on various seafloors with each seafloor's characteristics creating unique difficulties for detection. Three different experiments are conducted to study anomaly and

target detection on this particular dataset. Various parameters were changed to observe the differences in performance and to determine the best settings for detection. Template-based methods are compared against statistical methods. In the first experiment, K -Means versus K -Medoids estimated templates are studied. Experiment two compares the performance of combined frequency detection where the high and low frequency images are used together. The final experiment uses a much larger dataset to perform detection. The results of this experiment are compared to those of the smaller experiment. They show that template-based detection produces comparable results to statistics-based detectors. The spectral matched filter and Reed-Xiaoli detectors show the most promising results for anomaly and target detection when low frequency data is used.

CHAPTER 2

LITERATURE REVIEW

Chapter 2 starts with a brief overview of SONAR as it relates to synthetic aperture SONAR. This overview highlights the key elements of SONAR imagery that makes it useful for underwater imaging applications and consequently underwater object detection. Next, important factors and limitations that plague anomaly and target detection algorithms are discussed followed by a brief analysis of popular object detection algorithms whose successes have inspired subsequent works that improve upon their discoveries.

2.1 Overview of SONAR and Synthetic Aperture SONAR (SAS) Imagery

SONAR, which stands for sound navigation ranging, is a widely used technology for navigating and creating views of underwater scenes. Direct observations of underwater environments and the seafloor are not feasible due to poor visibility. Propagating sound waves can communicate with an underwater environment by returning echoes that bounce off of the environment and objects within. These ping echoes are then processed to form the scene that can be represented as an intensity image where the intensities are directly related to the amplitude of the echoed sound wave. SONAR is particularly useful underwater due to the fact that sound waves travel farther in water than light waves or radar. At about 1,500 meters per second, sound travels almost 5 times faster in seawater than in air with small variations from place to place [11]. Environmental conditions such as water temperature, salinity, pressure and depth can affect the speed of sound and vary the attenuation of a traveling sound wave. For these reasons, SONAR is used to map seafloors, develop nautical charts, locate underwater hazards to navigation or ordnance, and search for objects [12].

The National Oceanic and Atmospheric Association (NOAA) classifies SONAR into two types: active and passive. In active SONAR systems, acoustic signals or pulses of sound are emitted and bounce off objects in the path of the signals. These echoes are returned to transducers where the strength and time between emission and reception is measured to determine the range and orientation of the object [12]. On the other hand, a passive SONAR

system is primarily used to detect marine objects, such as submarines, ships and marine animals like whales [12]. These systems do not emit signals but instead detect noise or stray signals that could possibly come from marine objects [12]. Since passive SONAR systems do not emit output and are essentially “listening” for input signals, they cannot measure the range of objects.

There are various methods used to capture underwater SONAR imagery. Some involve the use of an autonomous underwater vehicle (AUV) that house the SONAR system and travels through the water. Others simply tow the system behind a vessel at a distance far enough so that noise from the vessel will not be captured. Images can be captured using forward-looking, upward-looking, or side-scan SONAR technologies, for example. In forward-looking SONAR the system is attached to the front of the vessel emitting forward propagating signals. It is particularly useful for navigation in shallow waters, ensuring a higher degree of safety [13]. Upward-looking SONAR systems operate by being moored at the subsurface. This technology is often used to monitor ice thickness and velocities, ocean current profiles and other in-situ water properties [14]. In side-scan SONAR systems, signals are emitted downward and at an angle from a towed vehicle. This technology is able to capture a wider area of the seafloor to the port (left) and starboard (right) sides of the system hence it's usefulness in object detection and seafloor mapping.

Additionally, most SONAR systems function at both low and high frequencies. Low frequencies tend to penetrate deeper into the seafloor making it suitable for buried object detection. High frequencies again reflect easier, effectively increasing detail in the processed image. Along with the differences in quality and detail of low and high frequency operation, it is important to note the range limitations of the system. Lower frequencies tend to travel farther and attenuate less in water which increases the range of the images [15]. Higher frequencies tends to have shorter range and attenuate more.

However, in this review, the focus will be on synthetic aperture SONAR (SAS) imagery and its usage in performing underwater anomaly and target detection. When large area

coverage and very high resolution is needed at the same time, SAS provides this capability. SAS is an example of an active SONAR system. It is an enhancement of standard SONAR technology that involves complex post-processing techniques which result in a higher resolution intensity image of an underwater scene. Traditional SONAR methods process each ping echo independently whereas SAS techniques process multiple broad-band ping echoes. Current systems are constructed with a single transmitter that emits acoustic pings along with an array of hydrophones and receivers that capture the ping echoes for processing [16]. The narrow beam-width pings are emitted as the SAS system travels creating a larger aperture. The coherent addition and processing of the echo pings by multiple hydrophones and receivers results in an intensity image with much higher resolution than standard SONAR. These images are usually more clear and crisp since aggregated ping echoes provide more information. The information from multiple pings goes into each pixel of the SAS image [17].

2.2 Motivation for Underwater Object Detection

High resolution SAS is especially useful for underwater object detection as it produces high quality images of underwater scenes. Using only sound waves at various frequencies, SAS is able to penetrate certain materials and develop clear depictions of dynamic underwater environments. In particular, underwater mine countermeasures (MCM) makes use of these high resolution images to detect and eliminate dangerous objects. This area of study focuses on safely and effectively locating potentially dangerous objects laying on or buried beneath the seafloor [9]. Often placed with malicious intent, objects such as mines can cause serious damage to nearby submarines, military or civilian ships and even marine life if they are detonated. Contrarily, an underwater anomaly detection method proposed in [18] used side-scan SONAR imagery to successfully detect wreckage from Air France Flight 447 that tragically crashed into the Atlantic Ocean in June of 2009. It follows that while underwater object detection serves an important purpose for safety, its usefulness spans as far as search and rescue missions to simply surveying our vast beautiful oceans for scientific research.

Over the years, interest in fully autonomous underwater detection systems has grown significantly. These systems are motivated by a desire to reduce human proximity to potentially harmful mines. Autonomous, or unmanned, underwater vehicles (AUV/UUV) have been manufactured to house SONAR or SAS systems and travel through water to capture images of these targets without endangering the safety of an onboard human operator. Pushing the autonomous capabilities even further, the added ability to process these images and determine whether an object is an anomaly or even a target is the new goal of many underwater detection research studies. Many target detection techniques employ anomaly detection as a first step in order to reduce computational costs and increase accuracy. In the following sections, important factors and limitations for successful and efficient detection along with popular anomaly and target detection methods noted in the literature are discussed.

2.3 Detection Limitations

In Williams [19, 20], several important limitations of the most popular detection algorithms are listed. These limitations were noted because most techniques either neglect or fail to address at least one of them. First, most detection algorithms assume uniform image quality across the entire image [19]. This is impractical in real life scenarios as the SONAR ping degrades at long ranges introducing noise into the images. Stability of the AUV system traveling through the water also contributes to poor quality. Second, the range-dependent nature of the highlight-shadow pattern of objects isn't sufficiently exploited as it can be predicted based on physics-based propagation models and geometrical considerations [20]. The authors mention that the signal-to-noise ratio (SNR) is weaker at shorter ranges, where the grazing angle of seafloor is closer to normal, and the length of shadows cast by an object is longer at longer ranges. Using this information, an expectation of the highlight-shadow pattern can be inferred.

Third, environmental conditions are improperly handled in supervised techniques where testing and training locations may be different [20]. Physical properties of the seafloor along with the frequency of the pings affect the outcome of the SAS image in unique ways. Hardness

and smoothness of the seafloor defines the amount of detail present in the image. If the seafloor has a hard, rocky bottom it will reflect most of the ping whereas a soft, muddy seafloor would absorb most of the ping or energy [15]. Hence, a hard rocky seafloor would be more detailed than a soft muddy seafloor. A smooth seafloor would act like a mirror and reflect sound in only one direction meaning the reflection wouldn't return to the SONAR system [15]. A rough seafloor with varied textures will have scattered echoes that return to the system. Therefore, care must be taken when using training data from one location that is applied to testing data in another that may have significantly different seafloor characteristics. One area of the seafloor could be composed of hard packed sand while the area immediately next to it could potentially be sand ripple or littered with rocks, biasing the detection results [19]. For instance, ripple detection is implemented in [20] as a way to reduce false alarms since they produce highlight-shadow patterns that can blend in with potential targets or be mistaken as objects.

The fourth noted limitation addresses the ambiguity of the detection threshold which is essentially an arbitrary score that has no physical meaning or definite value. However, if it is to be appropriately set by the detection operator, a principled way of determining it is needed. Finally, many detection algorithms require the full SONAR image to be available, preventing the possibility of "streaming" real-time detection [19]. The ability to perform detection on pieces of the image as they are scanned is particularly advantageous since detection can then be done in real-time on location. Due to these limitations, the problem of target detection is often preceded by anomaly detection. Anomaly detection attempts to bypass these limitations for target detection by removing irrelevant regions from the image. The goal of anomaly detection is to essentially filter the image to contain only regions of interest the target detection step.

Other issues that plague target detection involve characterization or recognition of the anomalies. Distinguishing anomalies from objects commonly found in underwater scenes is another research question. These issues are typically addressed when using prior knowledge

about the target to be detected making these methods supervised. A recognition method can even function as a target detector by excluding anomalies that are simply not targets.

2.4 Statistics- and Image Processing-Based Methods

Statistics-based methods often employ windows that calculate the mean or variance of pixels within a region of a SAS image in order to detect anomalies and targets [19], [20], [9]. These methods make assumptions about object highlight-shadow patterns. The characteristics of a pattern's pixel intensity values are used to compare local regions of the SAS image and determine if an anomaly is present. Thanks to integral image implementations [21], these methods are able to compute and compare region statistics relatively quickly.

One of the most popular statistical algorithms used in anomaly detection is the Reed-Xiaoli (RX) algorithm. It is commonly used as a prescreener for target detection and recognition in various modalities [9], [22]. For example, it has been used to successfully detect anomalies in hyperspectral imagery [23]. The RX algorithm achieves optimal detection performance in SAS imagery by simply locating pixels that are distinct from its surroundings [2]. In this sense, it functions as a unsupervised detector by using the statistics surrounding a center pixel to detect anomalies within the SAS image. It exploits the fact that objects appear in SAS imagery as highlight-shadow patterns surrounded by a majority of darker background pixels. The detector is described by a window composed of an inner and outer kernel that together create an annulus [2].

$$RX = \frac{(\mu_{in} - \mu_{out})^2}{\sigma_{out}^2} \quad (2-1)$$

The mean and variance of the kernels are calculated in order to compute a detector value. The inner and outer kernels are also separated by a predefined number of pixels. It is important that the inner kernel is large enough to completely encapsulate any potential anomaly so that background statistics are not contaminated with anomaly statistics. The output of the RX detector is defined by equation 2-1. The variables μ_{in} and μ_{out} represent the means of the pixel intensity values of the pixels in the inner and outer kernel, respectively. σ_{out}^2 is the variance of the outer kernel pixels. This detector essentially quantifies the difference between

the pixels in the regions of the two kernels. If the detector's output is high, the regions differ greatly whereas low values indicate that the two regions have little difference [1].

While the previously mentioned algorithms in Williams [19, 20] make use of properties of physics and background estimation to perform underwater object detection that overcome these limitations, they also rely heavily on extensive domain-specific knowledge about the problem. This kind of information may not always be readily available in practice, especially in anomaly detection. Anomaly detection methods often function in an unsupervised manner severely limiting the amount of prior knowledge. However, these algorithms employ a cascaded architecture that utilizes integral images to quickly generate background then shadow estimation maps based on image statistics in near real-time as the SAS image is produced. Inspired by the RX detector, background maps are computed locally around each pixel using a split-window technique. The window is split into three smaller horizontal windows centered on a pixel. The upper and lower horizontal windows are averaged to account for variations in seafloor characteristics of the local region surrounding the centered pixel. The shadow map computation overcomes range issues by using range-dependent window lengths to similarly calculate the region mean within the middle horizontal window. These maps allow the algorithms to adapt to range issues and seafloor characteristic changes so that regions of interest can be inferred.

Pixel intensity values in a shadowed region are lower than those in a background region since little to no echo is captured in the SAS image. For this reason, it is assumed that if a shadow is present, there is a possibility that an anomaly exists nearby. Consequently, a binary map highlighting the shadows is created which significantly reduces the final search area to only regions of interest. Clearly, these methods are capable of detecting underwater anomalies within SAS imagery while overcoming several detection limitations. The resulting binary maps are then used in subsequent target detection steps. From these methods, the importance of background and shadow statistic estimations are shown to be vital components to reduce the limitations of an accurate and efficient target detection method.

In Kaeli [18], image processing techniques are used to detect anomalies based on salience and rarity of objects in a scale-invariant fashion. Inspired by the scale-invariant feature transform (SIFT) [24], it aims to account for non-uniformity in seafloor depth. Non-uniformity poses issues with surveying underwater scenes since objects appear larger or smaller depending on the depth of the seafloor or alternatively, its distance from the SONAR system. Additionally, Kaeli [18] notes that at one scale a rock may appear anomalous until you decrease the image scale (i.e. zoom out) and notice that there are many other rocks in the scene removing the need to label it as anomalous. However, if you continue to zoom out, it is possible for the rock to be apart of a larger anomalous group of rocks surrounded by a flat uniform sandy seafloor. This requires an object to stand out from its immediate surroundings at multiple scales, confirming its salience. After determining if an object is salient, its rarity is observed to conclude if the object is unlikely to be found elsewhere hence, an anomaly. While this method is successful at locating any salient and rare object, it is unable to locate specific objects of interest or targets. For this reason, its purpose is better suited for determining if an anomaly is not present in an underwater scene. This can significantly reduce the search area in target detection applications, decreasing both computation and human intervention.

CHAPTER 3 METHOD

In this chapter, a description of the methodology employed for anomaly and target detection in this thesis is given. A supervised and an unsupervised approach is used. In the supervised approach, templates of targets within the SAS imagery are estimated and used along with information derived from the data to scan for anomalies and targets. Ideally, they are pictorial representations of potential targets found in the SAS dataset which share the same characteristics as actual targets such as the highlight-shadow pattern. The templates are used with various detector methods to individually produce a confidence value for every pixel location in an image. Together these pixel confidence values create a map for each template, slightly smaller than the original image, where homogeneous regions signify the presence or absence of an anomaly or target. For the unsupervised approach, detector methods which employ filters alone are used to create confidence maps. The confidence maps are further processed for final detection. The following detection steps are the same for both high and low frequency SAS images with the only difference being the input/output dimensions of the images and confidence maps. The following sections detail the steps used for supervised and unsupervised detection.

3.1 Estimate Templates

In this thesis, supervised detection techniques are utilized that rely on target templates estimated from training images within the dataset. The number of target templates estimated varies from four to eight with each template being used to search for anomalies. Two different versions of the templates are estimated for testing by taking advantage of the properties of objects appearing in SAS images. Provided with the dataset are the ground truth coordinates of all of the targets. These targets are extracted from their images and are used for training. Figure 3-1 shows the dimensions of the window containing the ground truth target which is extracted. Here, this extracted window containing the ground truth target is called a mugshot. It is rectangular in shape to accommodate the highlight-shadow pattern of an object

appearing in side-scan SAS. Ideally, the target is on the left of the window with its shadow being elongated and cast to the right. The left-centered green pixel in the figure denotes the location of the ground truth coordinate in the window. What follows is the process for template estimation.

Two popular clustering methods are used to estimate target templates: *K*-Means and *K*-Medoids. The objective of *K*-Means clustering is to partition mugshots into *K* clusters where each mugshot, defined as an image clip of an actual target from the dataset, is assigned to the closest cluster mean. It is an iterative algorithm that depends heavily on initial cluster means that are chosen at random. In each iteration, the cluster mean is updated by recalculating the cluster mean with the newly assigned mugshot. *K*-Medoids clustering works in a similar fashion. However, instead of assigning mugshots to the closest cluster mean, a single mugshot is initialized as the cluster medoid at random. In each iteration, the cluster medoid is reevaluated by calculating the average dissimilarity between the medoid and every mugshot within a cluster. The medoid is updated to be the mugshot with the smallest average dissimilarity meaning that it is the most “middle” mugshot in the cluster. The resulting cluster means and medoids are designated as the target templates. Template estimation however, is only used for supervised detection.

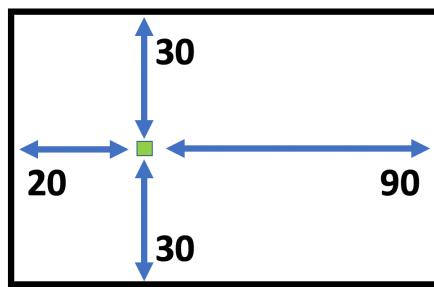


Figure 3-1. Dimensions of a mugshot, target template and detection window

Cross-validation is used with both *K*-Means and *K*-Medoids to estimate templates. After the templates are estimated, they are pruned to smooth jagged pixelated edges around the

target and to isolate the shadow and echo parts of the template. This is done using template pixel statistics and morphological operations commonly used in image processing.

This pruning step creates an additional target template that is used for testing where only the isolated shadow and echo pixels are used. Surrounding pixels are assigned the value of 0. In total, there are $2 \times K$ target templates used for testing: original template and pruned template. An example of this pruning process is shown in Figure 3-2 with the two final testing templates outlined in green.

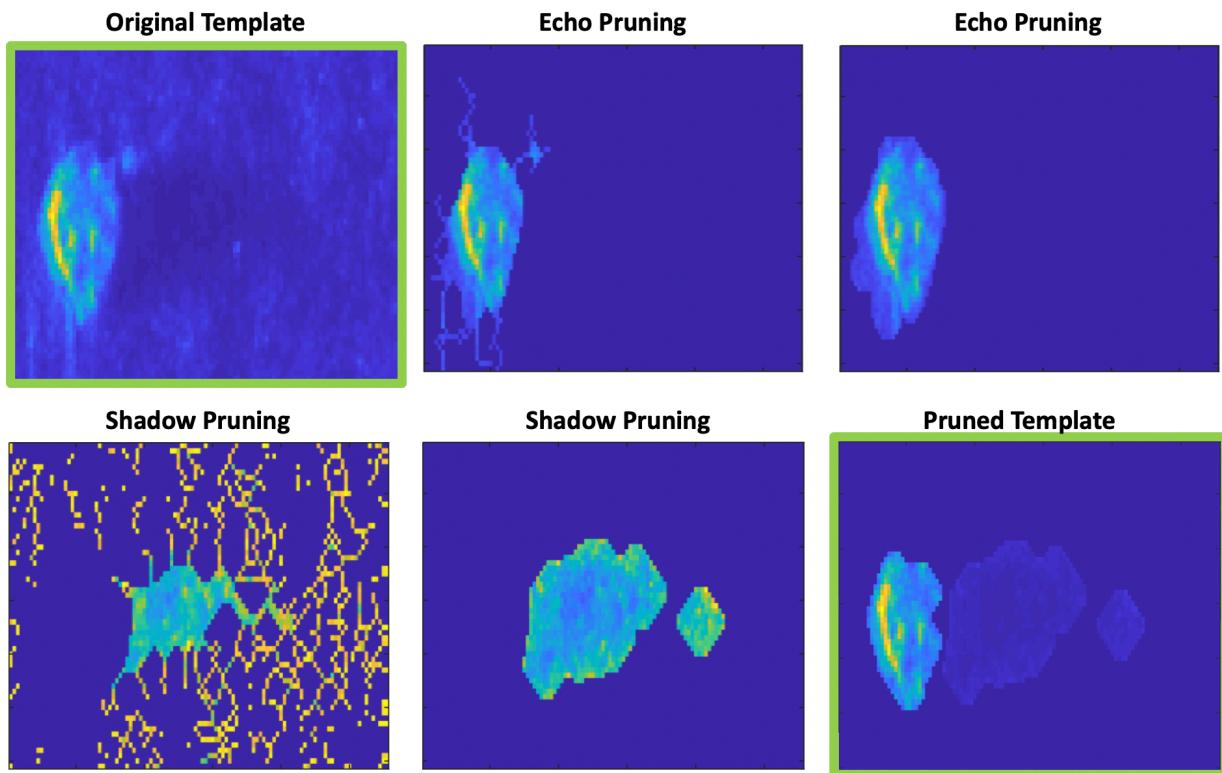


Figure 3-2. Example of template pruning process

3.2 Detectors: Produce Confidence Maps

To produce confidence maps, each estimated target template is used as a filter and is slid across the SAS image in a sliding window fashion, one pixel at a time. As each template slides across the image it is centered on a pixel, overlapping a window of the SAS image with the same size as the template. Confidence values are then calculated for each pixel using the

overlapping window and template as input to detector methods mentioned later in this section. Since the values are calculated at each pixel within the SAS image, a map is produced that is the same size as the original image. However, it is important to note that the resulting confidence map contains an annulus where pixels in the SAS image are ignored as the entire template does not fit within the image's boundaries. Figure 3-3 shows the diagram of this process.

What follows in this section are the various detector methods used to produce confidence maps along with their respective equations. Matched filters, adaptive cosine estimators, Reed-Xiaoli, convolution and normalized cross-correlation detectors are used. The Reed-Xiaoli detectors produce only one confidence map using no templates while the remaining detectors produce K confidence maps per image for both the K original and K pruned templates estimated using K -Means and K -Medoids. This means a total of $K \times 2 \times 2$ confidence maps are created for each image per detector. For each detector, let \mathbf{x} be the vectorized (i.e., reshaped into a vector) detection window. The detection window contains the enclosed pixels surrounding the centered testing pixel as shown in Figure 3-1. Let \mathbf{t} be a known vectorized target template. The operator $\|\cdot\|$ in the following equations represents the L_2 norm.

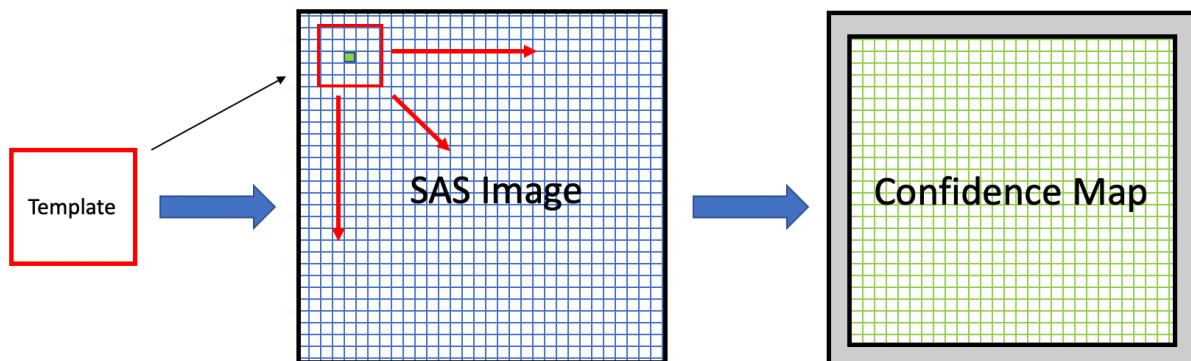


Figure 3-3. Diagram of how confidence maps are produced

3.2.1 Matched Filters

Matched filters (MF) are more commonly used in signal process applications to detect the presence of a target within a given signal. As images are just 2D representations of a signal, MF have also shown its usefulness in image processing applications. Since a target signal is known a priori when using MF, it is viewed as a supervised target detector. Equation 3–1 shows the calculation of a MF detector confidence value. The interval of values for this detector is $[0, \infty)$.

$$D_{\text{MF}}(\mathbf{x}, \mathbf{t}) = \frac{\mathbf{t}^T \mathbf{x}}{\|\mathbf{t}\|} \quad (3-1)$$

A notable variant of the MF is the spectral matched filter (SMF) which has performed well in sub-pixel target detection in hyperspectral imagery [25]. Again, this technique is commonly used in 1D signal processing but has been used here for 2D images. Equation 3–2 shows the calculation of a SMF detector confidence value where μ_b and Σ_b are the image background mean and covariance, respectively. Since most SAS images are mainly composed of background and targets take up a minimal amount of the image area, μ_b and Σ_b can be estimated using the sample mean and variance of the pixel intensities of each individual SAS image. The interval of values for this detector is $[0, \infty)$.

$$D_{\text{SMF}}(\mathbf{x}, \mathbf{t}) = \frac{(\mathbf{t} - \mu_b)^T \Sigma_b^{-1} (\mathbf{x} - \mu_b)}{\|\Sigma_b^{-1/2}(\mathbf{t} - \mu_b)\|} \quad (3-2)$$

3.2.2 Adaptive Cosine Estimators

The Adaptive Cosine Estimator (ACE) is a version of the SMF that computes the inner product between the target signature and a test sample in a whitened coordinate space [25]. The main difference between the two is that ACE normalizes \mathbf{x} before performing calculations. This makes the detector scale invariant so magnitude of pixels in the detection window do not influence the detector value. Equation 3–3 shows the calculation of the ACE detector confidence value where μ_b and Σ_b are the image background mean and covariance,

respectively.

$$D_{\text{ACE}}(\mathbf{x}, \mathbf{t}) = \frac{(\mathbf{t} - \mu_b)^T \boldsymbol{\Sigma}_b^{-1} (\mathbf{x} - \mu_b)}{\| \boldsymbol{\Sigma}_b^{-1/2}(\mathbf{t} - \mu_b) \| \| \boldsymbol{\Sigma}_b^{-1/2}(\mathbf{x} - \mu_b) \|} \quad (3-3)$$

The angle between the whitened versions of the test pixel \mathbf{x} and known target template \mathbf{t} is computed and returned as the detector confidence value on the interval $[-1, 1]$. High confidence values correspond to a small angle between the test pixel and target template indicating strong similarity. Low confidence values correspond to a large angle and dissimilarity.

Additionally, the Squared ACE detector, which is essentially just the square of the ACE detector, is used. The result of the ACE detector is squared to remove potential negative values and changing the value interval to $[0, 1]$.

$$D_{\text{ACE}^2}(\mathbf{x}, \mathbf{t}) = [D_{\text{ACE}}(\mathbf{x}, \mathbf{t})]^2 \quad (3-4)$$

3.2.3 Convolution

The 2D version of the convolution (CONV) is a direct extension of the 1D version though it convolves in both the horizontal and vertical directions. In image processing, convolutions are typically used for image smoothing, blurring, or edge detection with specifically designed filters. Here, CONV is used as a detector where the result of the calculated convolution between the detection window and template is considered a confidence value of the similarity between the two. Equation 3–5 shows the calculation of the detector where $\hat{\mathbf{t}}$ is the template vector after it has been reversed. It is equivalent to the MF without the denominator. The interval of values for this detector is $[0, \infty)$.

$$D_{\text{CONV}}(\mathbf{x}, \hat{\mathbf{t}}) = \hat{\mathbf{t}}^T \mathbf{x} \quad (3-5)$$

3.2.4 Normalized Cross-Correlation

Cross-correlation operators share similarities with the convolution. In image applications, cross-correlation is used for template matching. This makes it a suitable detector for finding a known pattern (target) in an image. It is another measure of the similarity between the detection window and template. Here, the normalized cross-correlation (NORMXCORR) is

used. Equation 3–6 shows the calculation of the detector where again, μ_b is the background mean and α_t is the mean of the template. NORMXCORR is similar to ACE but without the covariance. The interval of values for this detector is $[-1, 1]$.

$$D_{\text{NORMXCORR}}(\mathbf{x}, \mathbf{t}) = \frac{(\mathbf{x} - \mu_b)^T (\mathbf{t} - \alpha_t)}{\|\mathbf{x} - \mu_b\| \|\mathbf{t} - \alpha_t\|} \quad (3-6)$$

3.2.5 Reed-Xiaoli Detector

Described in Section 2.4, the RX detector is an unsupervised detector since it does not use prior information, i.e. templates. Instead this method relies on the statistics of the pixels within the detection window defined by the detector. A visual of the RX detection window is shown in Figure 3-4A with dimensions 41×41 pixels for the inner kernel and $201 \times 201 \oplus 151 \times 151$ pixels for the outer kernel (\oplus , symmetric difference) [9]. All of the dimensions are square. The output of the RX detector is again defined by Equation 2–1.

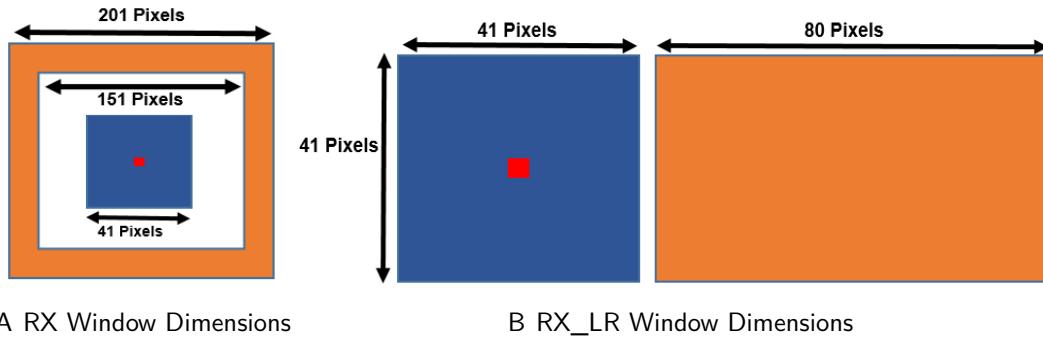


Figure 3-4. Kernel Dimensions for RX Methods [1]

It is recognized that objects present in the SAS imagery appear as highlight-shadow patterns surrounded by a majority of dark background pixels. The shadow tends to fall on the far side of the SAS camera. The RX Left-Right (RX_LR) detector [9] exploits this characteristic of side-scan SAS imagery by altering the detection window of the original RX detector. The outer kernel of the RX detector is moved to use the statistics of the region of pixels to the right of the object. These pixels are usually darker because of the shadow. The new kernel is now a rectangle that surrounds the shadow cast by the object. This allows the

detector to capture the entire highlight-shadow pattern, thus the name. Equation 3-7 defines the calculation of the RX_LR detector where μ_{left} is the mean of the left kernel, μ_{right} is the mean of the right kernel and σ_{right}^2 is the variance of the right kernel. Figure 3-4B shows the kernel for the RX_LR detector along with its dimensions. The interval of values for this detector is $[0, \infty)$.

$$D_{RX_LR}(\mathbf{x}) = \frac{(\mu_{left} - \mu_{right})^2}{\sigma_{right}^2}. \quad (3-7)$$

3.3 Combine Confidence Maps

Once an image has K number of confidences maps made with a particular detector, using the K templates, they are then combined to create a final map used for detection. By combining the maps, variability of the estimated templates is accounted for in their individual resulting confidence maps. These K confidence maps are combined by point-wise pixel multiplication or addition of the maps. This results in one combined confidence map per detector for the image. This is done for every image in the dataset. Figure 3-5 shows this process.

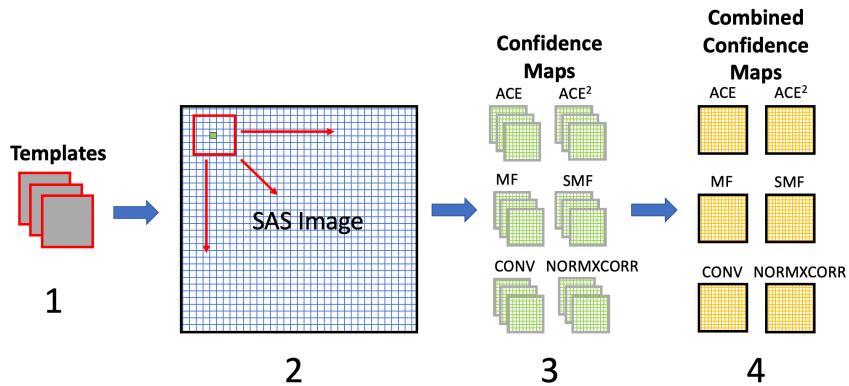


Figure 3-5. Example of combining confidence maps for $K = 3$ with either K -Means or K -Medoids original or pruned templates

A pessimistic approach is taken by multiplying the confidence maps. It is essentially a binary AND operation where all of the K confidences maps need to detect an anomaly or target in order for it to be present in the combined confidence map. For instance, if one

confidence map has a region of 0s where every other map has pixel intensity values greater than 0, the multiplication operation will have 0s in that region of the combined map. By adding the confidence maps, the idea is to strengthen the confidence values of the detectors. More trust is placed on the individual detectors with the assumption that they are able to confidently find strong similarities between the templates and targets in the images. This step is skipped for the RX detectors since multiple confidence maps are not produced.

3.4 Non-Maximum Suppression and Detection

The final step before detection is to perform non-maximum suppression on the combined confidence maps of the template-based detectors and the confidence maps of the statistics-based RX detectors to locate targets and reduce false alarms or anomalies. Non-max suppression locates the highest pixel intensity values, or peaks, in the map and suppresses pixels within a particular radius. This step however is done in two different ways. The first involves thresholding the images in the testing fold with the background mean of the training folds. This is done only with the supervised detectors. Basically, the average pixel intensity of all of the training images is calculated and any pixels below this value in the testing image is set to 0. The other way is by not using background thresholding at all. Figure 3-6 shows an example result of non-max suppression with and without background thresholding.

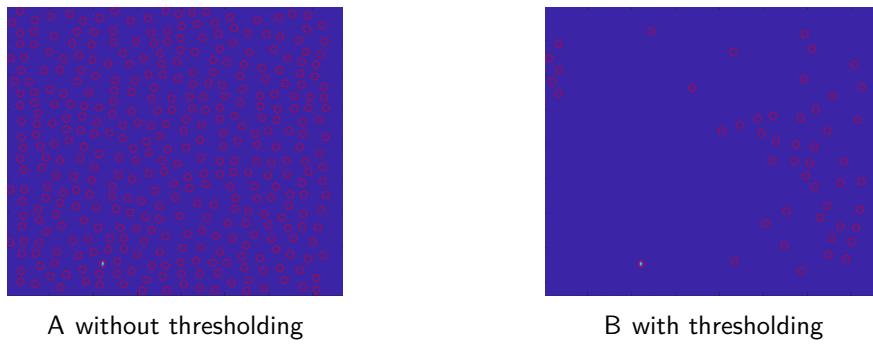


Figure 3-6. Example of non-maximum suppression with and without background thresholding

The peaks remaining after non-max suppression are detected anomalies. By using the target templates to scan the imagery for objects that are similar to the templates, the goal is a reduction in false alarms on anomalies while increasing confidence in detecting actual targets.

To quantify this goal, detection results are displayed using receiver operating characteristic (ROC) curves. On the x -axis is the false alarm rate (FAR) which ranges from 0 to 5 and on the y -axis is the true positive ratio (TPR) which is the number of true positives that are classified as such.

CHAPTER 4 EXPERIMENTS AND DISCUSSION

In this section, a description of the experimental setups are given along with results and discussion. The entire dataset contains 6,760 high and low frequency SAS images collected at five different sites exhibiting varying seafloor characteristics and containing varying targets. These sites are labeled as Site A-E. Numerous experiments were run on two different subsets of the data. The first subset contains 180 images from Site A. Not every image in this subset contains a target. The second subset includes 278 images from all five sites with each image containing at least one target. Details of the number of images and targets in each subset are shown in Tables 4-2 and 4-15.

The original images have been lanczos [26] down-sampled to reduce computation times and storage. The final dimensions of the images are 1680x1548 for high frequency images and 840x1548 for low frequency images. There are four different parameters that can be changed to vary the template-based experiments. Table 4-1 shows the options for the parameter settings. The experiments ran on each subset of the data is explained in the following sections. The template-based experiments are done using *K*-Means and *K*-Medoids estimated templates.

Table 4-1. Template-Based Experimental Setup Parameter Options

Template Type	Frequency	Number of Templates	Combine Map Method
Original	High	4	Addition
Pruned	Low	6	Multiplication
Combined			

4.1 *N*-Fold Cross-Validation

To remove bias within training (estimating templates), cross-validation is used by separating the images in such a way that no test image is involved in the training process. Each image is placed into a group and these groups are referred to as folds. Of the *N* folds

created, $N - 1$ is used for training with the single held out fold being used for testing. This process is done N times by rotating the the folds so that each fold is allowed to serve as a testing fold. There is no need for cross-validation for the statistics-based anomaly detectors used as there is no training. An example of this is shown in Figure 4-1

4.2 Site A*: K-Means vs K-Medoids

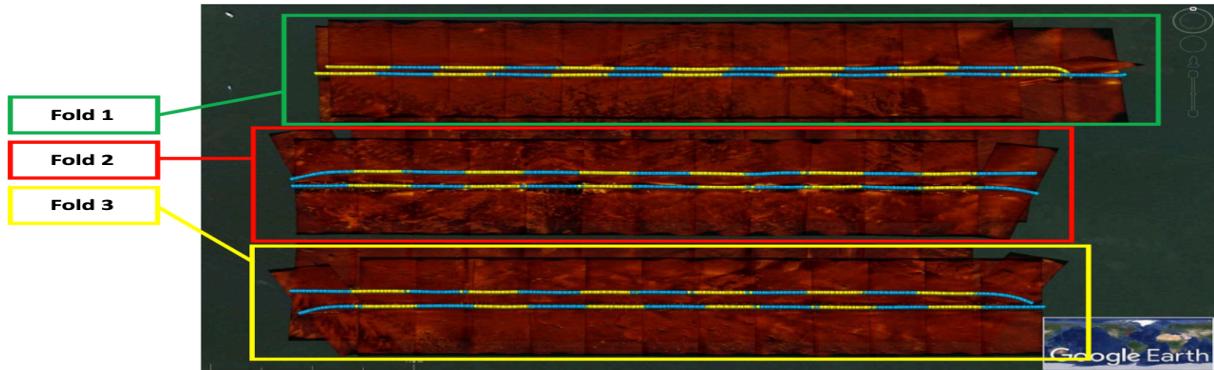


Figure 4-1. AUV path and fold designation

For the first experiment, 180 images from Site A are separated into three folds for cross-validation with each fold containing 60 images. This subset of Site A is designated as Site A*. There are 16 targets in this subset meaning that every image does not contain a target. Figure 4-1 shows how the folds are designated for cross-validation. They are chosen based on the path of the AUV capturing the SAS images. Each fold has the images captured by two paths of the AUV. The figure shows that none of the paths in a fold overlap with another. In general, Site A is characterized by its varied seafloor texture, many scattered non-target objects, and linear shaped targets. Site A* simulates a real-world scenario in surveying underwater scenes for MCM since the amount of captured area containing potential targets is usually very small compared to the amount of non-target area captured. Table 4-2 shows the Site A* breakdown of folds and the targets in each.

Table 4-2. Number of images and targets in each fold of Site A*

Site A*		
Fold	Number of Images	Number of Targets
1	60	4
2	60	5
3	60	7
Total	180	16

For both K -Means and K -Medoids, the number of templates estimated is four and six.

For each set of templates, a series of experiments are run using the method introduced in Chapter 3. The frequency of the templates and images is chosen and the type of template used for detection is varied between the original estimated template and pruned template. The method for combining the individual confidence maps is multiplication. Additionally, background thresholding is used in effort to reduce the number of false alarms. However, this also increases the chance of missing actual targets. The threshold is applied to the testing fold by computing the average pixel intensity of all images in the training folds. During non-maximum suppression, only peaks above the threshold are considered. Again, all of these options are shown in Table 4-1. Using these options, there are 12 different experiments that can be done on the Site A* subset. As an example, one experimental setting is as follows: number of templates is four, original templates are used, frequency of images/templates used is high, and the maps are combined using multiplication. Figure 4-2 shows a diagram of this example.

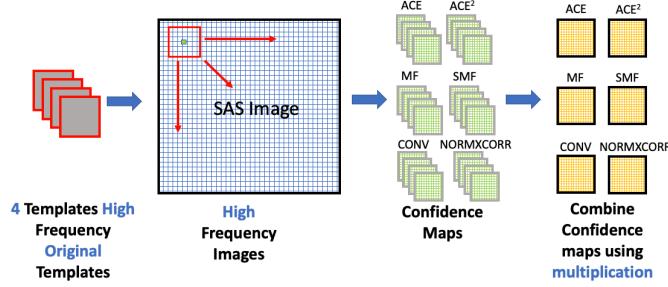


Figure 4-2. Example of an experiment with four original high frequency templates and high frequency images. The four individual confidence maps are combined using multiplication.

In Sections 4.2.1 and 4.2.2, the results of the various K -Means, K -Medoids, RX and RX_LR Site A* experiments are shown using ROC curves and a table of the TPRs for selected FARs. The results of each combination of target template method and frequency is compared with the statistic-based results. Section 4.2.1 shows the results using original templates and Section 4.2.2 shows the results using pruned templates. Note that the RX and RX_LR results do not change when changing the template type, but only when changing frequency. Figures of ROC curves comparing the TPR vs FAR for each detector is given. In each figure, the high frequency ROC curve results (left) are compared to the low frequency ROC curve results (right). For all experiments, the FAR is represented by a generic scale on the range [0,5] with lower FAR being highly desired for this MCM application. In each table, the entries are the TPR for each detector for both high (H) and low (L) frequency testing at the following FAR: 0.1, 0.3, 0.5, 1, 3 and 5. The number of templates used is given and for all template-based experiments and again, the confidence maps are combined using multiplication. Section 4.2.3 summarizes the findings of these experiments.

Note that the RX and RX_LR results are the exact same in each figure and table in this section. These detectors are statistics-based and do not rely on templates and prior information making them unsupervised. The frequency of the data is the only parameter change for these detectors.

4.2.1 Original Templates

Figure 4-3 shows the results of the K-Means experiments on Site A* using four original templates along with the statistics-based results. It can be seen that all of the detectors have different capacities for detecting targets at certain FARs. For high frequency data (4-3A), each template-based detector is able to detect all 16 targets with a FAR of less than or equal to 5, thus achieving a TPR of 1. In Table 4-3 this can be observed in the high frequency (H) column of each detector. For these columns the TPR for each detector increases to 1. SMF is clearly the best performing template-based detector achieving a TPR of 1 at 0.3 FAR. In contrast, for the same detector using the low frequency data (4-3B), the maximum TPR achieved is 0.875 meaning only 14 targets are able to be detected within the desired FAR range. There are two template-based detectors that suffer the most when using low frequency data: MF and CONV. The MF reaches a maximum TPR of 0.5 while CONV reaches 0.563. Their results are similar likely due to the fact the calculation of the detector values are similar.

The statistics-based RX detector actually has similar performance to the template-based detectors when using high frequency. Although, with low frequency it is the only detector to reach a TPR of 0.938 meaning that it detected 15 of the 16 targets. One more than the template-based methods. RX_LR has the poorest performance of all the detectors, finding only 9 and 6 targets for high and low frequency, respectively.

However, an important thing to note from the ROC curves is that for template-based detection when using low frequency data, most of the detectors are able to reach a higher TPR at lower FARs than high frequency data. This can be seen in the ROC curves as well as Table 4-3. When comparing the H and L column for the detectors at the lowest FAR, the TPR for low frequency is typically higher than that for high frequency.

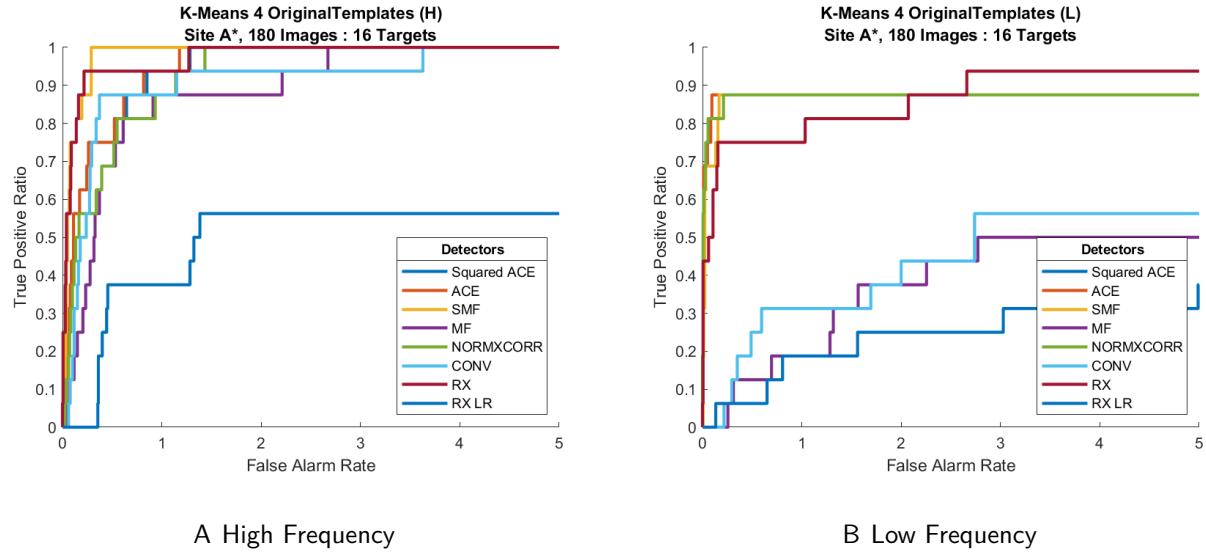


Figure 4-3. *K*-Means 4 Original Templates ROC Curves

Table 4-3. *K*-Means 4 Original Templates table of selected FAR vs TPR

False Alarm Rate	True Positive Ratio - 16 Targets															
	K-Means: 4 Original Templates								No Templates							
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		RX		RX_LR	
	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L
0.1	0.438	0.875	0.438	0.875	0.75	0.688	0.125	0	0.313	0.813	0.125	0	0.75	0.5	0	0
0.3	0.75	0.875	0.75	0.875	1	0.875	0.438	0.063	0.563	0.875	0.75	0.125	0.938	0.75	0	0.063
0.5	0.75	0.875	0.75	0.875	1	0.875	0.688	0.125	0.688	0.875	0.875	0.25	0.938	0.75	0.375	0.063
1	0.938	0.875	0.938	0.875	1	0.875	0.875	0.188	0.875	0.875	0.875	0.313	0.938	0.75	0.375	0.188
3	1	0.875	1	0.875	1	0.875	1	0.5	1	0.875	0.938	0.563	1	0.938	0.563	0.25
5	1	0.875	1	0.875	1	0.875	1	0.5	1	0.875	1	0.563	1	0.938	0.563	0.375

Figure 4-4 and Table 4-4 show the *K*-Medoids Site A* results using four original templates and the statistics-based results. These results are very similar to the previous *K*-Means results showing no drastic differences in performance though *K*-Means performs slightly better. SMF again showed the best performance for high frequency, achieving a TPR of 1 at 0.5 FAR. The remaining template-based detectors reach a TPR of 1. The *K*-Medoids ROC curves and table display the same trends as *K*-Means for both frequencies. In low frequency, MF and CONV also experience decreased performance detecting just half of the

targets while the other supervised detectors find 14. The RX and RX_LR results are the exact same as in the previous figure.

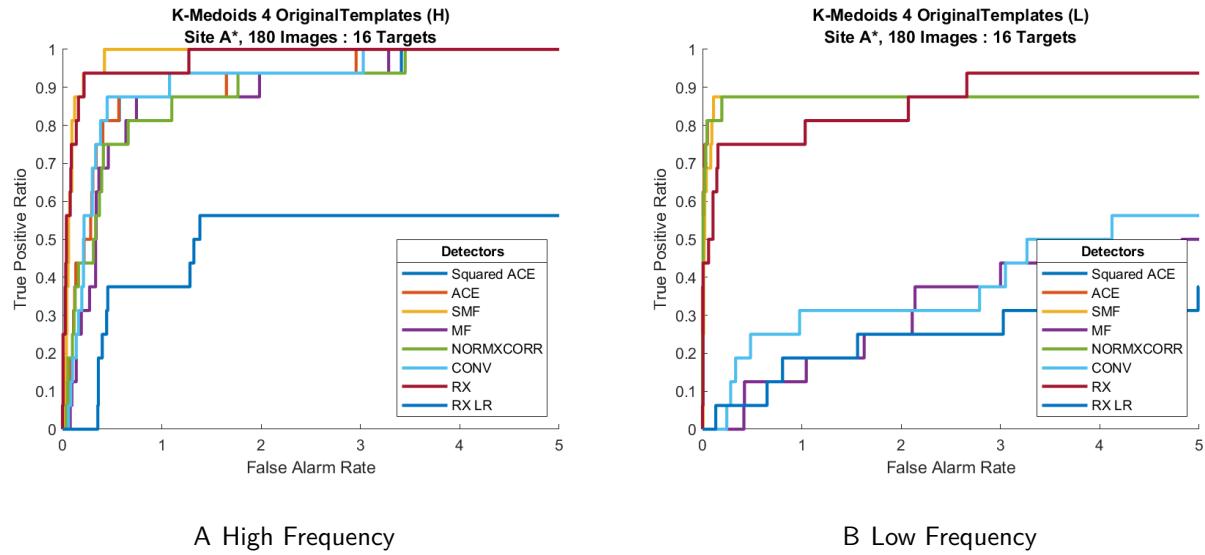


Figure 4-4. *K*-Medoids 4 Original Templates ROC Curves

Table 4-4. *K*-Medoids 4 Original Templates table of selected FAR vs TPR

False Alarm Rate	True Positive Ratio - 16 Targets															
	<i>K</i> -Medoids: 4 Original Templates															
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		RX		RX_LR	
H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	
0.1	0.188	0.813	0.188	0.813	0.813	0.813	0.125	0	0.188	0.813	0.125	0	0.75	0.5	0	0
0.3	0.625	0.875	0.625	0.875	0.938	0.875	0.375	0	0.438	0.875	0.563	0.125	0.938	0.75	0	0.063
0.5	0.813	0.875	0.813	0.875	1	0.875	0.75	0.125	0.75	0.875	0.875	0.25	0.938	0.75	0.375	0.063
1	0.875	0.875	0.875	0.875	1	0.875	0.875	0.125	0.813	0.875	0.875	0.313	0.938	0.75	0.375	0.188
3	1	0.875	0.938	0.875	1	0.875	0.938	0.375	0.938	0.875	0.938	0.375	1	0.938	0.563	0.25
5	1	0.875	1	0.875	1	0.875	1	0.5	1	0.875	1	0.563	1	0.938	0.563	0.375

For the next set of experiments, the number of templates was increased to six to see if the *K*-Means and *K*-Medoids algorithms were able to create more distinct target templates capable of capturing the variability of the targets. Figures 4-5 and 4-6 and Tables 4-5 and 4-6 show the ROC curves and TPR tables of the *K*-Means and *K*-Medoids Site A* experiments,

respectively. The RX and RX_LR results are again the same in the following tables since they are unsupervised detectors and do not use templates.

For K-Means, high frequency is not able to achieve a TPR of 1 for most template-based detectors using six target templates. SMF and MF are the only detectors able to detect all 16 targets for the desired maximum FAR other than RX. All other supervised detectors were able to detect 15 of the 16 targets at 5 FAR with a TPR of 0.938. This provides information about the six templates estimated by the K-Means algorithm. Evidently, six K-Means templates may be too many for detecting the 16 targets when compared to the four template experiments since the performance decreases. Again, the MF and CONV detectors have degraded performance using the low frequency data as shown in Figure 4-5B for K-Means. Higher TPRs are also observed at lower FAR for the other four detectors.

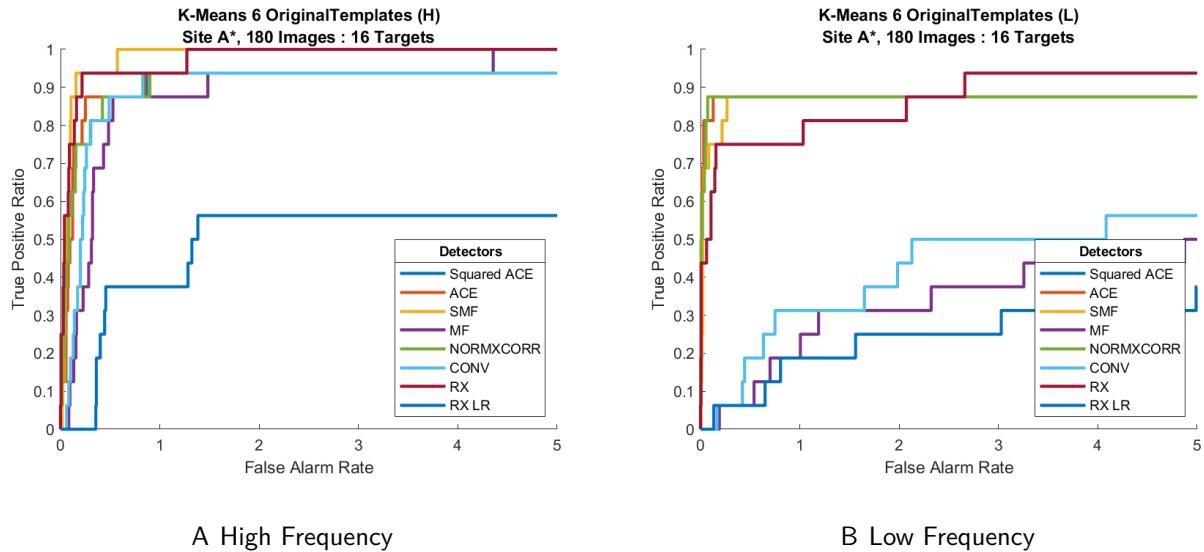


Figure 4-5. K-Means 6 Original Templates ROC Curves

Table 4-5. *K*-Means 6 Original Templates table of selected FAR vs TPR

False Alarm Rate	True Positive Ratio - 16 Targets																	
	K-Means: 6 Original Templates												No Templates					
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		RX		RX_LR			
	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L
0.1	0.5	0.813	0.5	0.813	0	0.75	0.125	0	0.563	0.875	0.125	0	0.75	0.5	0	0	0	
0.3	0.875	0.875	0.875	0.875	0	0.875	0.438	0.063	0.75	0.875	0.75	0.063	0.938	0.75	0	0.063		
0.5	0.875	0.875	0.875	0.875	0.75	0.875	0.813	0.063	0.875	0.875	0.875	0.188	0.938	0.75	0.375	0.063		
1	0.938	0.875	0.938	0.875	0.938	0.875	0.875	0.188	0.938	0.875	0.938	0.313	0.938	0.75	0.375	0.188		
3	0.938	0.875	0.938	0.875	0.938	0.875	0.938	0.375	0.938	0.875	0.938	0.5	1	0.938	0.563	0.25		
5	0.938	0.875	0.938	0.875	1	0.875	1	0.5	0.938	0.875	0.938	0.563	1	0.938	0.563	0.375		

K-Medoids has similar performance to *K*-Means when using six target templates for both frequencies. Ironically, it slightly outperforms *K*-Means using six target templates for both high and low frequency when comparing their respective TPR values. When comparing the TPR tables, *K*-Medoids generally reaches a higher TPR before *K*-Means in regards to increasing FAR. This is especially apparent with high frequency. With six templates the high frequency ACE, SMF and MF detectors are able to detect all of the targets for *K*-Medoids while only SMF and MF are able to do so for *K*-Means. The results for low frequency are pretty much the same for both template method.

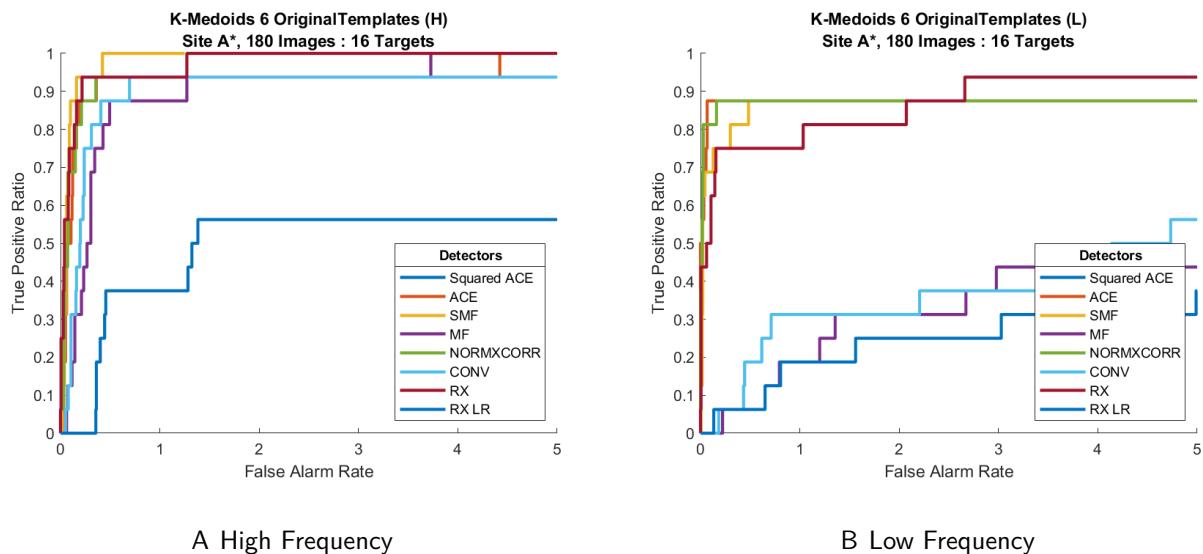


Figure 4-6. *K*-Medoids 6 Original Templates ROC Curves

Table 4-6. *K*-Medoids 6 Original Templates table of selected FAR vs TPR

False Alarm Rate	True Positive Ratio - 16 Targets																	
	K-Medoids: 6 Original Templates												No Templates					
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		RX		RX_LR			
	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L
0.1	0.5	0.875	0.5	0.875	0	0.688	0.125	0	0.688	0.813	0.125	0	0.75	0.5	0	0	0	
0.3	0.875	0.875	0.875	0.875	0	0.75	0.5	0.063	0.875	0.875	0.75	0.063	0.938	0.75	0	0.063		
0.5	0.938	0.875	0.938	0.875	0.813	0.875	0.875	0.063	0.938	0.875	0.875	0.188	0.938	0.75	0.375	0.063		
1	0.938	0.875	0.938	0.875	0.938	0.875	0.875	0.188	0.938	0.875	0.938	0.313	0.938	0.75	0.375	0.188		
3	0.938	0.875	0.938	0.875	1	0.875	0.938	0.438	0.938	0.875	0.938	0.375	1	0.938	0.563	0.25		
5	1	0.875	0.938	0.875	1	0.875	1	0.438	0.938	0.875	0.938	0.563	1	0.938	0.563	0.375		

The results of using four original templates show that the supervised SMF detector has the best overall performance with *K*-Means being the better template estimation method. For high frequency using four templates with both *K*-Means and *K*-Medoids templates, SMF is able to detect all 16 targets consistently. Both template estimation methods have similar performances for all detectors in low frequency with degraded performance for the MF and CONV detectors. For the other template-based detectors, 14 of the 16 targets are detected with a TPR of 0.875.

The overall performance decreases when using six original templates. However, with high frequency data and six templates the better performing template estimation method is *K*-Medoids. The high frequency *K*-Medoids ACE, SMF and MF detectors detect all targets whereas only *K*-Means' SMF and MF are able to. Low frequency results have minimal differences between *K*-Means and *K*-Medoids with degraded performances for MF and CONV again observed and the results are very similar to those when using four templates. It is likely that *K*-Medoids does better than *K*-Means with six templates because the algorithm uses the actual mugshots as templates instead of derived templates like *K*-Means. In essence, six *real* mugshots are used as templates to identify the objects as opposed to six averaged templates that are *similar* to the real mugshots. The six differing *K*-Medoid templates also compensate for added object variability as opposed to four. Other than this distinction, *K*-Means and *K*-Medoids exhibit similar performance trends as shown in the ROC curves. Again, for both

template methods and for four and six templates, detection is better at the lower FARs using low frequency.

4.2.2 Pruned Templates

Figure 4-7 shows the results of high and low frequency K-Means experiments on Site A* with four pruned templates using ROC curves. Again, it can be seen that all of the detectors have different capacities for detecting the targets at certain FAR when using pruned templates. For high frequency (4-7A), the SMF, MF, and NORMXCORR detectors are able to detect the targets with a TPR of 1 with a FAR of less than or equal to 5. CONV is able to reach 0.938, detecting 15 targets while ACE and ACE² are only able to reach 0.125 detecting only 2 for the same FAR. SMF outshines all other detectors with these settings as it achieves a TPR of 0.875 at a FAR of 0.1 whereas the average TPR at this FAR for all supervised and unsupervised detectors combined is 0.375.

For low frequency (4-7B) however, no detectors are capable of detecting all of the targets. ACE and ACE² are not able to detect any targets. SMF detects only 1 target, experiencing the largest reduction in performance compared to high frequency as it goes from being able to detect all targets at 0.3 FAR using high frequency to only 1 target at 5 FAR. MF, NORMXCORR, and CONV detect 14 targets compared to all 16 being detected using high frequency. Although, these particular detectors are more certain at lower FARs since at 0.1 FAR their TPRs are much higher than those of the high frequency detectors at the same FAR as noted in Table 4-7. For easy comparison, the RX and RX_LR results are included in the table. These results are the same here for both four and six templates as those in the previous experiment using the original templates.

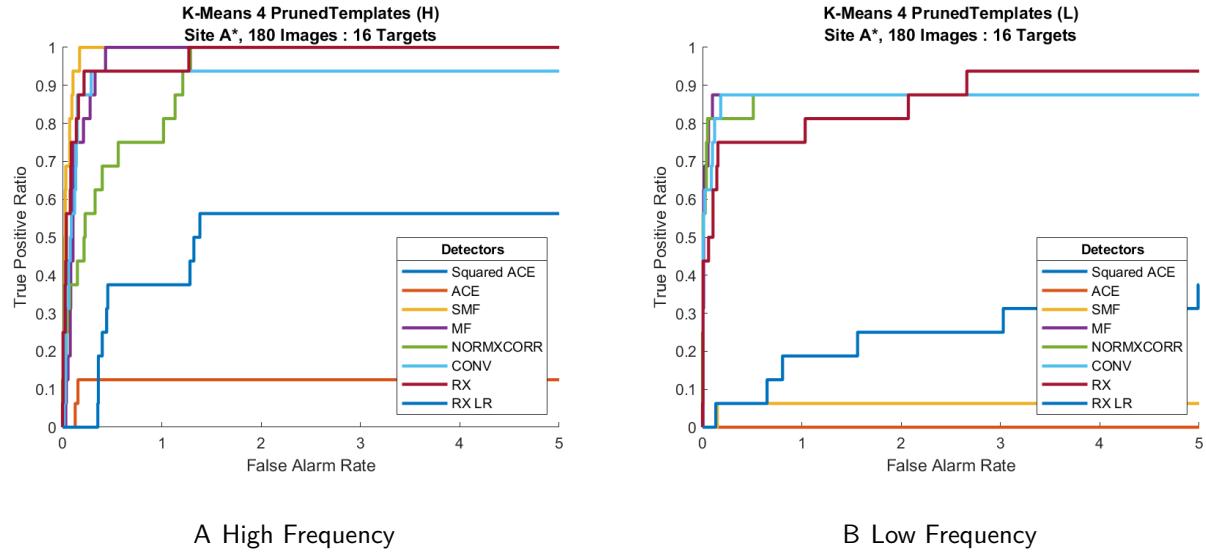


Figure 4-7. *K*-Means 4 Pruned Templates ROC Curves

Table 4-7. *K*-Means 4 Pruned Templates table of selected FAR vs TPR

False Alarm Rate	True Positive Ratio - 16 Targets															
	<i>K</i> -Means: 4 Pruned Templates															
	ACE								No Templates							
	H	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H
0.1	0	0	0	0	0.875	0	0.438	0.813	0.375	0.813	0.563	0.688	0.75	0.5	0	0
0.3	0.125	0	0.125	0	1	0.063	0.875	0.875	0.563	0.813	0.938	0.875	0.938	0.75	0	0.063
0.5	0.125	0	0.125	0	1	0.063	1	0.875	0.688	0.813	0.938	0.875	0.938	0.75	0.375	0.063
1	0.125	0	0.125	0	1	0.063	1	0.875	0.75	0.875	0.938	0.875	0.938	0.75	0.375	0.188
3	0.125	0	0.125	0	1	0.063	1	0.875	1	0.875	0.938	0.875	1	0.938	0.563	0.25
5	0.125	0	0.125	0	1	0.063	1	0.875	1	0.875	0.938	0.875	1	0.938	0.563	0.375

At first glance, Figure 4-8 and Table 4-8 shows that *K*-Medoids using four pruned templates does not perform as well as its *K*-Means counterpart. Interestingly, for high frequency, ACE and ACE² are now able to reach a TPR of 0.25 compared to *K*-Means' 0.125. SMF, MF and CONV are able to reach a TPR of 1 while NORMXCORR reaches 0.813. This result is similar to that of *K*-Means except that NORMXCORR is able not to reach a TPR of 1 though CONV does. For low frequency, none of the detectors are able to detect all of the targets. This is the same as with *K*-Means with ACE and ACE² detecting no targets in

this case. Again, K -Medoids detectors detect less targets than K -Means for the same FAR when using four templates. Higher TPRs are again observed at lower FAR for low frequency compared to high frequency.

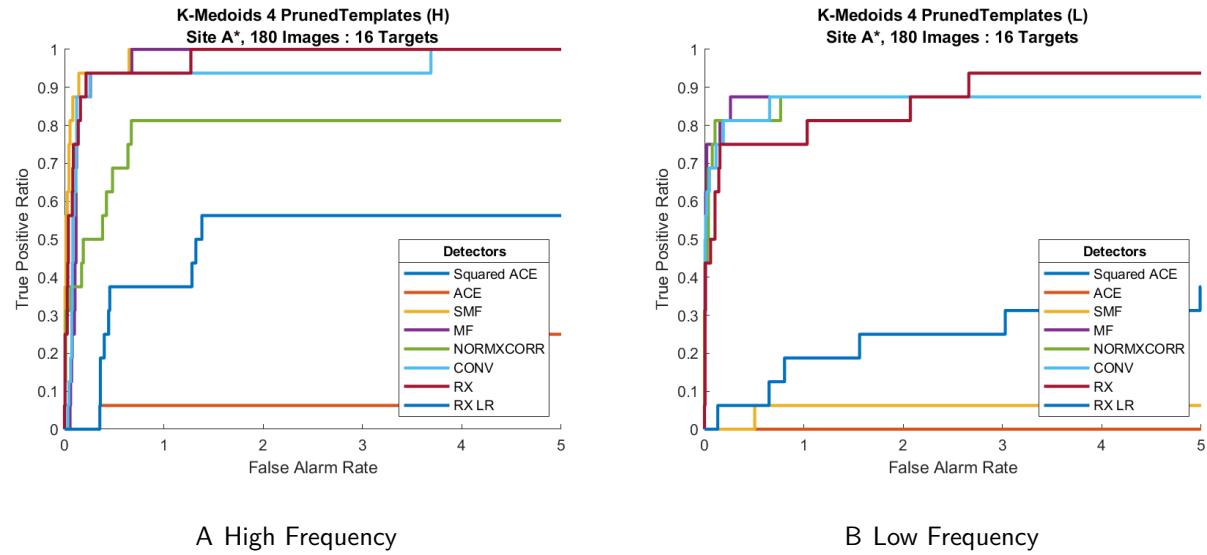


Figure 4-8. K -Medoids 4 Pruned Templates ROC Curves

Table 4-8. K -Medoids 4 Pruned Templates table of selected FAR vs TPR

False Alarm Rate	True Positive Ratio - 16 Targets															
	K -Medoids: 4 Pruned Templates															
	ACE								No Templates							
	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L
0.1	0	0	0	0	0.875	0	0.25	0.75	0.375	0.75	0.563	0.688	0.75	0.5	0	0
0.3	0	0	0	0	0.938	0	0.938	0.875	0.5	0.813	0.938	0.813	0.938	0.75	0	0.063
0.5	0.063	0	0.063	0	0.938	0	0.938	0.875	0.688	0.813	0.938	0.813	0.938	0.75	0.375	0.063
1	0.063	0	0.063	0	1	0.063	1	0.875	0.813	0.875	0.938	0.875	0.938	0.75	0.375	0.188
3	0.063	0	0.063	0	1	0.063	1	0.875	0.813	0.875	0.938	0.875	1	0.938	0.563	0.25
5	0.25	0	0.25	0	1	0.063	1	0.875	0.813	0.875	1	0.875	1	0.938	0.563	0.375

The number of templates was increased to six again for the pruned target templates. Figures 4-9 shows the ROC curves for these results. The performance for the high frequency ACE and ACE² detectors (4-9A) was the worst for K -Means achieving a TPR of 0.188 and 0.125, respectively. Besides these two detectors, all other supervised detectors are able to

achieve a TPR of at least 0.938 for high frequency. MF and CONV are the only detectors to reach a TPR of 1. low frequency (4-9B) performance degrades for all detectors. None are able to detect all targets while ACE and ACE² detect no targets. The best performing low frequency supervised detectors are MF, NORMXCORR and CONV all detecting 14 targets with a TPR of 0.875. The low frequency trend appears again in these results; higher TPRs are observed at lower FARs when compared to high frequency. Table 4-9 shows the TPRs.

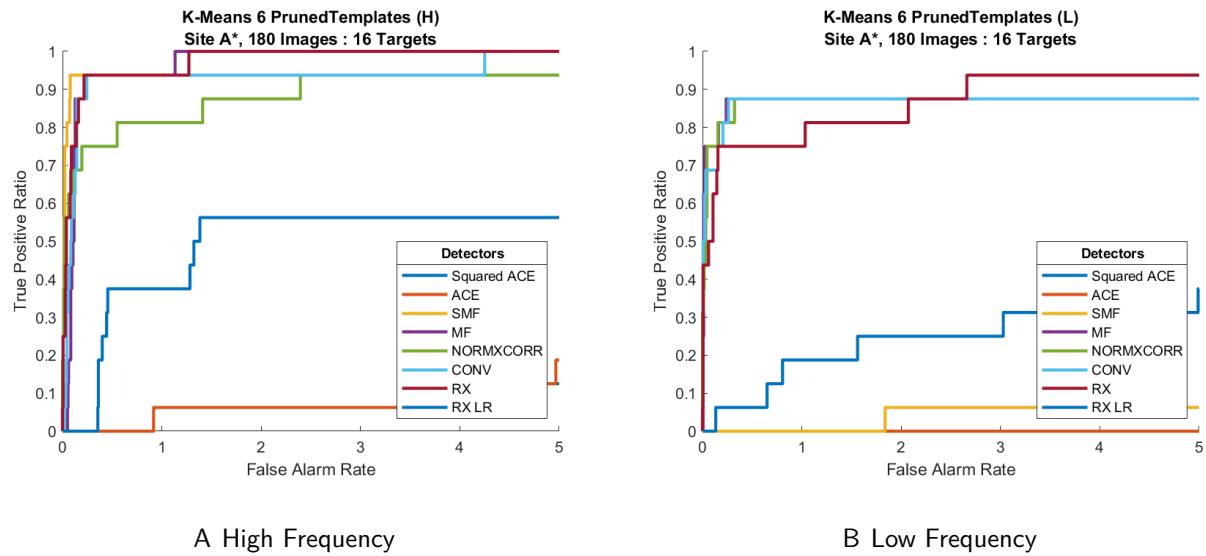


Figure 4-9. *K*-Means 6 Pruned Templates ROC Curves

Table 4-9. *K*-Means 6 Pruned Templates table of selected FAR vs TPR

False Alarm Rate	True Positive Ratio - 16 Targets															
	<i>K</i> -Means: 6 Pruned Templates															
	No Templates															
	ACE	ACE ²	SMF	MF	NORMXCORR	CONV	RX	RX_LR	H	L	H	L	H	L	H	L
0.1	0	0	0	0	0.938	0	0.438	0.75	0.625	0.75	0.563	0.688	0.75	0.5	0	0
0.3	0	0	0	0	0.938	0	0.938	0.875	0.75	0.813	0.938	0.875	0.938	0.75	0	0.063
0.5	0	0	0	0	0.938	0	0.938	0.875	0.75	0.875	0.938	0.875	0.938	0.75	0.375	0.063
1	0.063	0	0.063	0	0.938	0	0.938	0.875	0.813	0.875	0.938	0.875	0.938	0.75	0.375	0.188
3	0.063	0	0.063	0	0.938	0.063	1	0.875	0.938	0.875	0.938	0.875	1	0.938	0.563	0.25
5	0.188	0	0.125	0	0.938	0.063	1	0.875	0.938	0.875	1	0.875	1	0.938	0.563	0.375

The results of using six pruned K -Medoids templates for detection are given in Figure 4-10 ROC curves and Table 4-10. Only two detectors, SMF and MF, are able to detect all targets in high frequency as shown in subfigure 4-10A. As seen in the four pruned template results, a boost in performance occurs for high frequency with these two detectors when compared to the original templates. ACE and ACE² again has poor performance using pruned templates reaching a TPR of only 0.188. NORMXCORR and CONV both reach 0.938 which can be considered sufficient performance. Performance drops for low frequency with ACE and ACE² being the worst by detecting none of the targets and SMF detecting only 1 with a TPR of 0.063. From Figure 4-10B, the ROC curve shows that MF, NORMXCORR and CONV all have similar performances reaching the same TPR of 0.875 and detecting 14 targets.

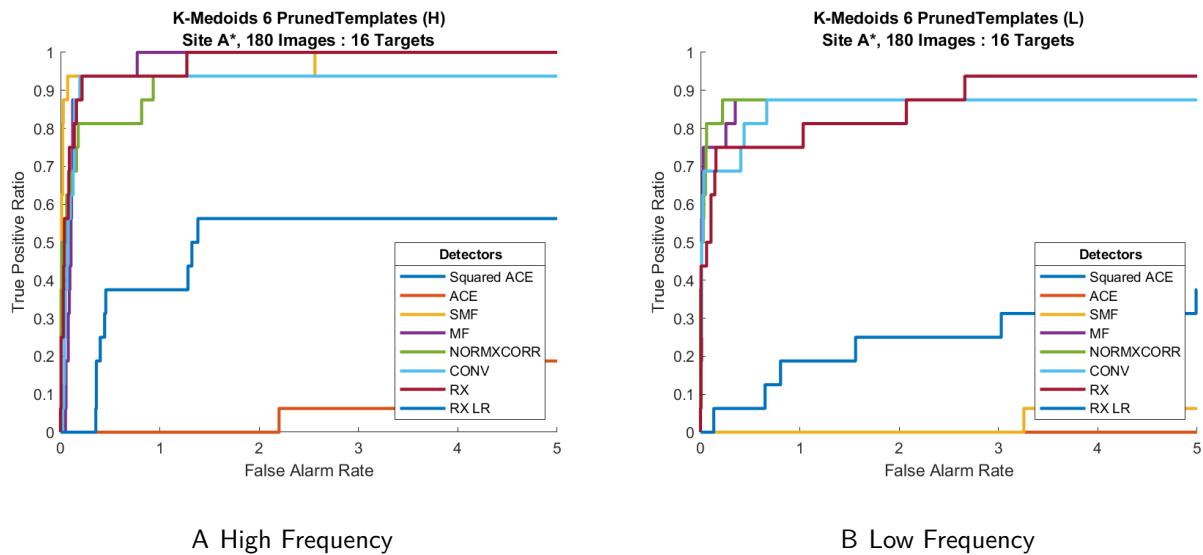


Figure 4-10. K -Medoids 6 Pruned Templates ROC Curves

Table 4-10. K -Medoids 6 Pruned Templates table of selected FAR vs TPR

False Alarm Rate	True Positive Ratio - 16 Targets																	
	K -Medoids: 6 Pruned Templates												No Templates					
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		RX		RX_LR			
	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L
0.1	0	0	0	0	0.938	0	0.438	0.75	0.625	0.813	0.563	0.688	0.75	0.5	0	0	0	0
0.3	0	0	0	0	0.938	0	0.938	0.813	0.813	0.875	0.938	0.688	0.938	0.75	0	0.063		
0.5	0	0	0	0	0.938	0	0.938	0.875	0.813	0.875	0.938	0.813	0.938	0.75	0.375	0.063		
1	0	0	0	0	0.938	0	1	0.875	0.938	0.875	0.938	0.875	0.938	0.75	0.375	0.188		
3	0.063	0	0.063	0	1	0	1	0.875	0.938	0.875	0.938	0.875	1	0.938	0.563	0.25		
5	0.188	0	0.188	0	1	0.063	1	0.875	0.938	0.875	0.938	0.875	1	0.938	0.563	0.375		

The result of using four and six pruned K -Means and K -Medoids templates are given above. For the high frequency K -Means experiment, four templates seems to out perform six with more detectors able to detect all of the targets typically at lower FARs. The results using low frequency are essentially identical for both numbers of templates. The results of using four and six K -Medoids templates are very similar. The only significant boost in detection is by SMF and MF when using high frequency data. The remaining template-based detectors have minimal differences in performance by switching this parameter. Additionally, no significant difference in performance is observed by switching from low to high frequency. For each experiment in this section, once again the TPRs for the low frequency results are lower than those for high frequency at the lower FARs.

4.2.3 Summary

Sections 4.2.1 and 4.2.2 give the results of using four and six original and pruned templates along with K -Means and K -Medoids estimated templates for detection. Each section discusses the performance of both template estimation methods. The ROC curves for these experiments are a quick visualization of the performance and trends can be easily observed by comparing the original and pruned figures.

The performance of the detectors with high frequency exhibit similar trends across parameter configurations. SMF is consistently one of the best performing detectors. SMF and

MF improve for four and six templates when pruned templates are used versus when original templates are used. On the other hand, the remaining supervised detectors' performance degrades when using pruned templates and high frequency with either four or six templates. The ACE and ACE² detectors suffer the most, detecting no more than 4 targets with a TPR of 0.25. Additionally, no significant changes occur when using *K*-Means versus *K*-Medoids estimated templates. The unsupervised RX detector uses no templates and has its best performance with high frequency data. In some instances the results for RX are about the same or better than the best supervised detector, SMF. On the other hand, the RX_LR detector performs better than the ACE detectors but worse than every other detector including RX.

The performance of all of the detectors decreases when low frequency data is used. Each of the low frequency ROC curves have the same trend when comparing the performance of the detectors for both *K*-Means and *K*-Medoids with no significant differences between the two template methods. NORMXCORR is the only detector with consistently good performance for low frequency using either original or pruned templates and four or six templates. It shows only slightly better detection using original over pruned templates. MF and CONV are the only two template-based detectors with poor performance using the original templates. With pruned templates in general, SMF accompanies ACE and ACE² as the worst performing detectors while the other supervised detectors are able to reach TPRs of 0.875 with either four or six templates. The RX and RX_LR detectors experienced a decline in performance just as the template-based detectors. The RX detector has the better performance of the two with, consistently having equal or higher TPRs than the best performing template-based detector.

Overall, the best configuration for supervised template-based detection is by using the original templates and high frequency data. For both *K*-Means and *K*-Medoids all of the detectors have sufficient performance and similar results with neither being a standout. Although, four templates work better at detecting all 16 targets. If pruned templates are used with high frequency, NORMXCORR, ACE and ACE² detectors suffer. Four templates again

proves to be the best number for detection. For these reasons, the a system using the SMF detector with four original K -Means templates and high frequency data would be preferable for supervised detection. However, the unsupervised high frequency RX detector would be most preferable since it generally matches or outperforms all others in every experiment by detecting all targets without parameter tuning. This experiment also showed that it that low frequency data is actually better at detecting more targets at much lower FARS. Depending on the specifications of the system, this frequency may be chosen for application.

4.3 Site A*: Combined Frequency

For this experiment, the results of combining high and low frequency maps for detection is studied. The objective here is to see if combined frequency confidence maps can improved detection. To do this, the methodology introduced in Section 3 for generating confidence maps with four and six K -Means and K -Medoids templates are used for the template-based detectors. This results in a high and low frequency confidence map generated per SAS image. The main difference for this experiment is that these two confidence maps are combined to create one combined frequency confidence map that is then used for detection. The idea is to increase the magnitude of the confidence values corresponding to anomalies. This is done using the Site A* dataset. Note that there are only four variations of the RX and RX_LR results and are the same for all experiments in this section since no templates are used. Therefore, the RX results are shown in figures and tables throughout the section and discussed in Section 4.3.3.

Since the dimensions of the high and low frequency maps are different, two versions of the final combined frequency map are created to match either the high or low frequency map dimensions. To combine the high and low frequency maps of a SAS image, either the high or low map is resized using bi-cubic interpolation to match the dimensions of the other. Next, the maps are either added or multiplied together to produce the final combined frequency confidence map. This combination of variables results in four different combined frequency confidence maps that are used for detection, producing four ROC curves per template method.

On the x-axis is the arbitrary FAR ranging from 0 to 5 with values closer to 0 being desired.

On the y -axis is the TPR ranging from 0 to 1 with values closer to 1 being desired.

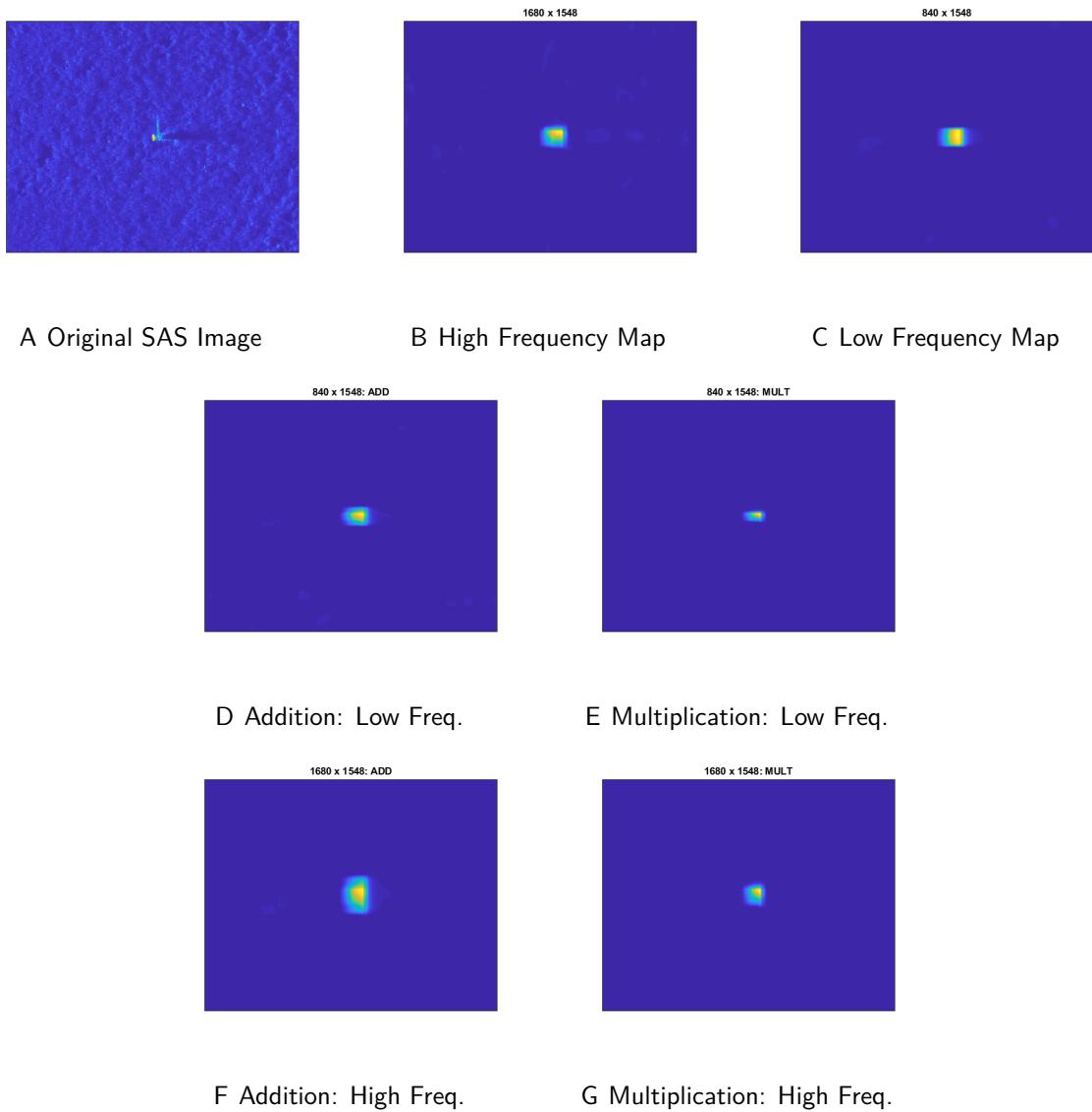


Figure 4-11. Cropped images of the combined confidence maps and combined frequency
Confidence Maps

Adding the maps ideally increases the magnitude of the pixel confidence values.

Background regions are assumed to have lower values so although the region's confidence values will be higher, anomalies should still be distinguishable. Anomalies are assumed to have higher values so the pixels representing the anomaly would have much higher confidence

values than surrounding background regions. Multiplying the maps is ideally a targeted way of increasing the magnitude of the confidence values of homogeneous regions corresponding to anomalies and decreasing the values of regions corresponding to background. Following the aforementioned assumptions, multiplying large values would return large values and multiplying smaller values would return small values. This method of combining the maps using addition is similar to a logical OR operation whereas multiplication is similar to a logical AND operation.

Figure 4-11 shows an example of the combined confidence maps and combined frequency confidence maps using a clipped portion of a SAS image.

4.3.1 Four Original Templates

In this section the combined frequency results using four original templates are shown using ROC curves. For each template method, four ROC curve figures are shown. The ROC curves corresponding to trials where addition is used to combine the high and low frequency confidence maps are shown above those where multiplication is used to combine the maps. The left ROC curves correspond to the combined frequency maps where high frequency dimensions are used and the right corresponds to where low frequency dimensions are used. The TPR tables accompany the ROC curves for detailed analysis.

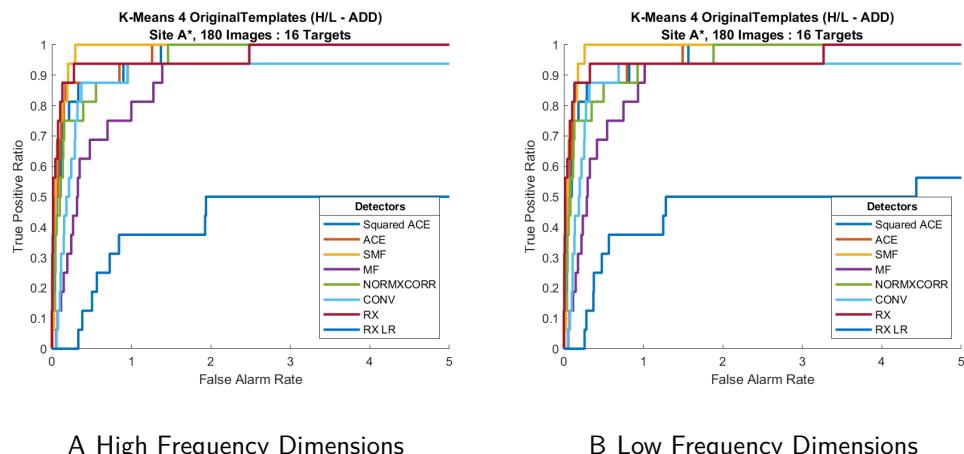


Figure 4-12. *K*-Means Addition Combined Frequency 4 Original Templates ROC Curves

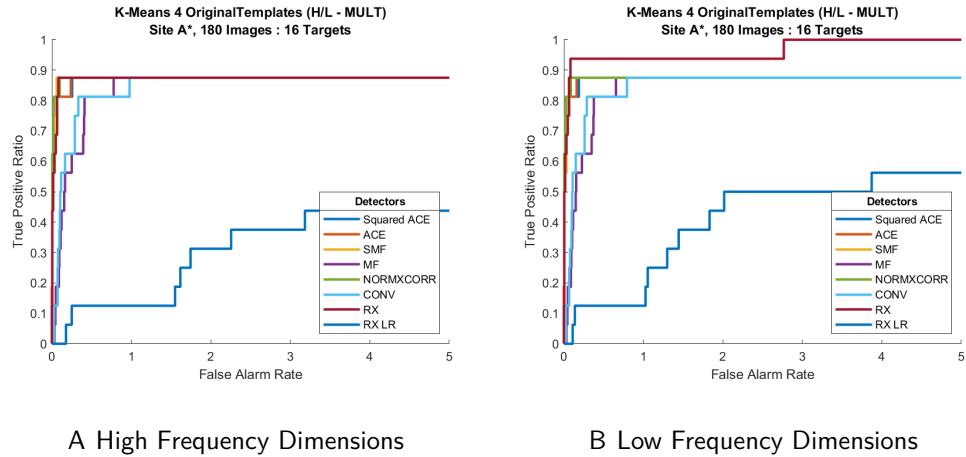


Figure 4-13. *K*-Means Multiplication Combined Frequency 4 Original Templates ROC Curves

The results shown above are from using four *K*-Means original templates. By observing these ROC curves, it is clear that addition is the best way of combining these confidence maps. For every detector, supervised or unsupervised, addition has the best results. For Figures 4-12A and 4-12B, the template-based detectors are able to detect the targets within the desired FAR range for both frequency dimensions with the exception of the MF and CONV detectors that reached a TPR of 0.938 missing only one target. Just as the experiments in Section 4.2, higher TPRs are observed, for lower FAR using low frequency dimensions as opposed to high frequency dimensions in some cases. SMF has another standout performance using the combined frequency confidence maps. It is the first detector to achieve the highest TPR at a FAR of 0.1 among all of the template-based detectors. The rest of the detectors have nearly identical performances for both frequencies when adding the maps.

Figures 4-13A and 4-13B show the result of this experiment using multiplication to combine each individual frequency confidence maps. The main distinction between multiplication and addition is that none of the supervised template-based detectors are able to detect all 16 targets within the desired FAR range. All of these detectors reach a maximum TPR of 0.875 finding only 14 targets. Higher TPRs are again observed with lower

FARs using the low frequency dimensions. Here, SMF is able to find the 14 targets at a FAR of 0.1, sooner than any other supervised detector.

A trend that occurs in Table 4-11 and tables to follow is that at certain FARs there are no TPR values. At these FARs the detector has achieved its maximum TPR. For the cells with no values, there are actually no more peaks greater than the ROC threshold that corresponds the particular FAR and TPR. Only the peaks that remain after the non-maximum suppression step in the experimental method (Subsection 3.4) are available to generate a TPR. Thus, there is no way for the TPR to increase so the FAR remains constant. This threshold value varies for each detector within the detector's value range. This is one of the main reasons FAR and TPR are used to compare the detectors instead of a threshold which is actually a chosen confidence value for an individual detector.

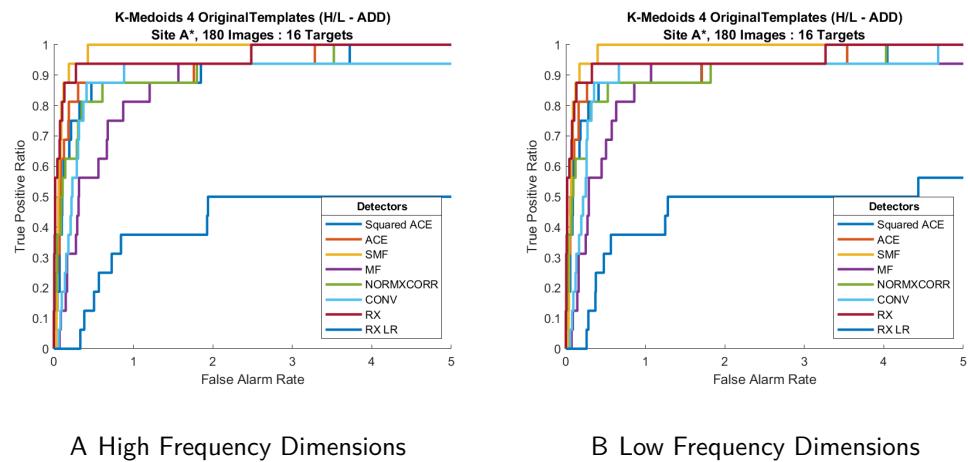


Figure 4-14. *K*-Medoids Addition Combined Frequency 4 Original Templates ROC Curves

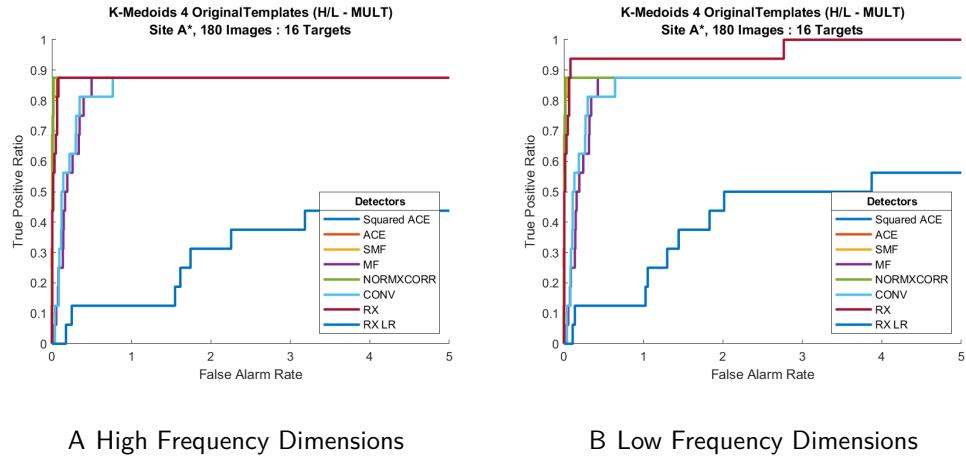


Figure 4-15. *K*-Medoids Multiplication Combined Frequency 4 Original Templates ROC Curves

Figures 4-14, 4-15 and Table 4-12 present the results of the combined frequency confidence maps using four original *K*-Medoids templates. These results are very similar to those using the *K*-Means templates. The differences in these four trials are not clear in the ROC curves and can be better observed by comparing the TPR values in Table 4-12. When adding the maps, SMF again has the best performance by detecting all 16 targets within the desired FAR range for both frequencies. The CONV detector however is now able to detect all of the targets when using the low frequency dimensions for the combined frequency confidence map with *K*-Medoids templates. Both the high and low frequency dimension MF detectors are able to only detect 15 targets along with the high frequency dimension CONV detector. When using multiplication to produce the combined frequency confidence maps, the same trend exhibited with *K*-Means occurs here. All of the detectors reach the same TPR of 0.875, detecting 14 targets.

Overall, the supervised method of using four original *K*-Means or *K*-Medoids templates works well to detect the 16 targets using the combined frequency confidence maps. Using addition to create the combined map is preferable as many of these template-based detectors are able to detect all of the targets. Multiplication only allows 14 of the 16 targets to be detected but with higher TPRs at lower FARs. This applies when comparing the respective

frequency dimensions. The K -Means four original template combined frequency results can also be compared to the K -Means four original template individual frequency results in Section 4.2.1. There is some improvement for both template methods when using combined frequency confidence maps with high frequency dimensions for either adding or multiplying. The low frequency combined confidence maps show the most improvement. At the lowest FAR, using the combined frequency confidence maps increases performance for each supervised detector. Combined frequency provides a significant boost in performance for both the CONV and MF detectors using addition and multiplication with low frequency dimensions.

4.3.2 Six Original Templates

In this section, six original K -Means and K -Medoids templates are used to create combined frequency confidence maps for supervised detection. Four ROC curves are produced by varying the dimensions of the combined frequency confidence maps and the method used to combine them.

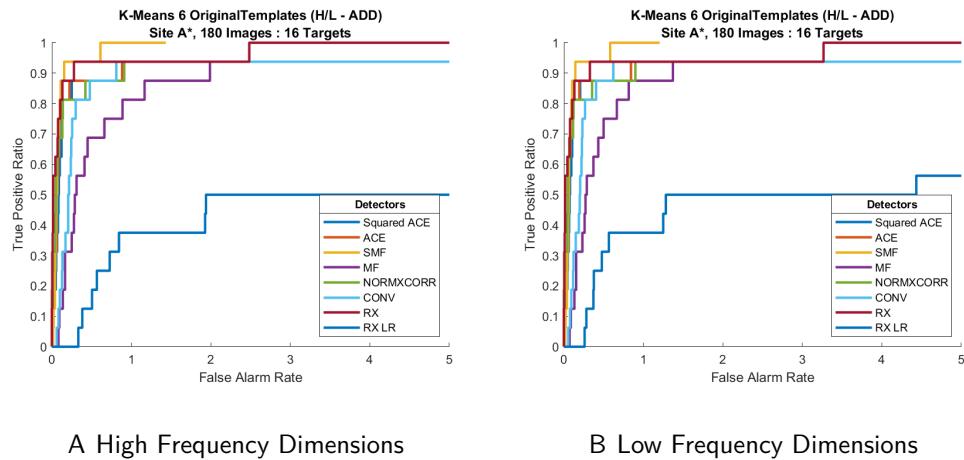


Figure 4-16. K -Means Addition Combined Frequency 6 Original Templates ROC Curves

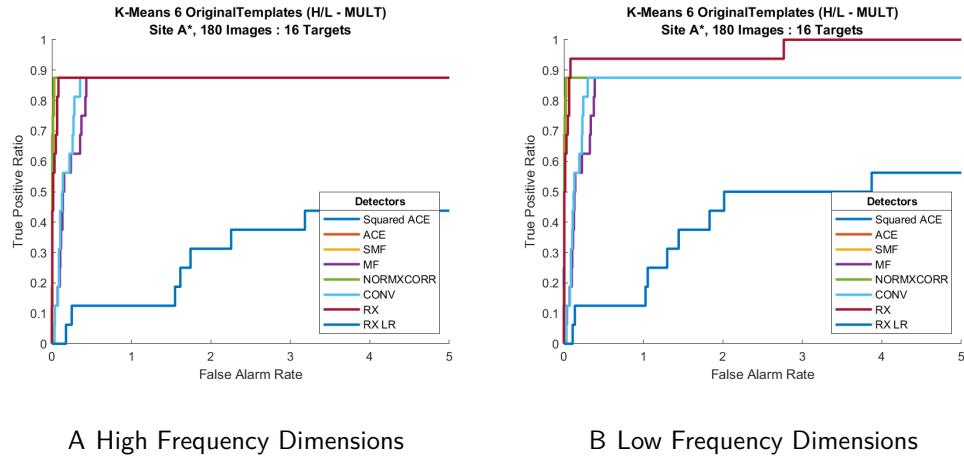


Figure 4-17. *K*-Means Multiplication Combined Frequency 6 Original Templates ROC Curves

Figure 4-16 shows the *K*-Means ROC curves when addition is used to combine the confidence maps. The confidence maps either have the dimensions of the high frequency SAS images or the low frequency SAS images. Figure 4-17 shows the ROC curves when multiplication is used. For addition and multiplication, the ACE, ACE², SMF and MF detectors the results are pretty much the same when comparing both dimensions. SMF had the best performance when addition is used, being the only detector to find all 16 targets while the rest of the detectors reached a TPR of 0.938 and detecting 15 targets. The take-away from this experiment is that addition works better when using the combined frequency confidence maps with six *K*-Means templates for detection. For every template-based detector, the final TPR values in Table 4-13 for addition are higher than the corresponding TPR values for multiplication. Although for the lowest FAR of 0.1, multiplication has higher TPRs than addition for these supervised detectors.

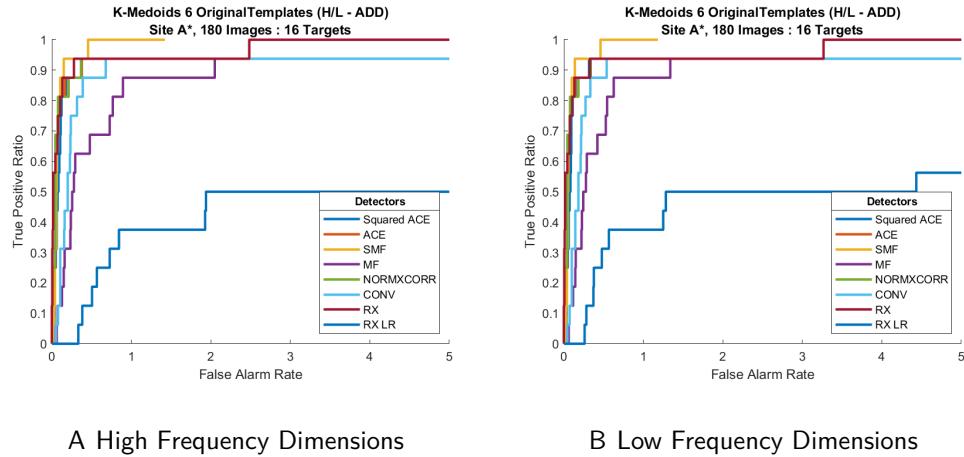


Figure 4-18. *K*-Medoids Addition Combined Frequency 6 Original Templates ROC Curves

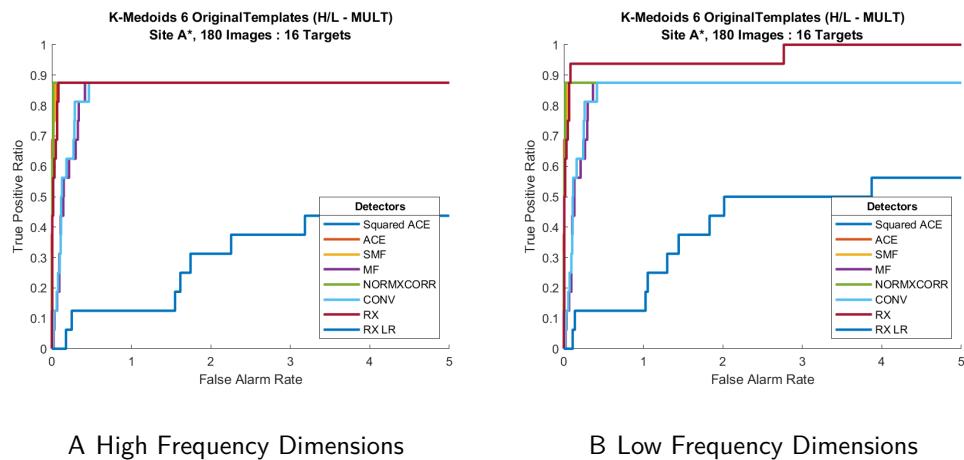


Figure 4-19. *K*-Medoids Multiplication Combined Frequency 6 Original Templates ROC Curves

Figure 4-18 and 4-19 shows the supervised detection ROC curves using combined frequency confidence maps from six *K*-Medoids templates for addition and multiplication, respectively. The results coincide with the *K*-Means detection results. For addition and multiplication, the TPRs are essentially the same when changing the dimensions of the combined frequency confidence maps. SMF again is the best performing detector being the only one to reach a TPR of 1 within the desired FAR range. Again, this is done using addition to combine the maps. The rest of the detectors reach a maximum TPR of 0.938 for addition.

None of the template-based detectors are able to detect all 16 targets using multiplication thus they all reach a maximum TPR of 0.875.

4.3.3 Summary

According to the results of the experiments in Sections 4.3.1 and 4.3.2 the dimensions of the combined frequency confidence maps matters when a preference in desired FAR is specified. For most of the experiments, the TPR values for the combined frequency confidence maps with high frequency dimensions and combined frequency confidence maps with low frequency dimensions are essentially identical. Addition is the best method for combining the high and low frequency confidence maps as it is the only way to detect all 16 targets. Although, multiplication is best to use when a lower FAR is desired since this setting achieved higher TPRs. With addition, SMF is the best detector to use in this instance as it has success detecting more targets at lower FARs than the other detectors.

Finally, of the two unsupervised statistics-based detectors, when adding the maps RX has the better performance. It reaches a TPR of 1 with both frequencies. RX_LR is unable to detect any targets at a FAR of 0.1 while RX is able to reach a TPR of 0.75 for both frequency dimensions. Eventually, RX detects all of the targets while RX_LR detects about half of the targets with either frequency dimension. However, the RX detector is the only supervised or unsupervised detector able to reach a TPR of 1 while using multiplication to combine the frequency confidence maps of all the variations of detectors shown in this section. It also reaches a TPR of 0.938 at a FAR of only 0.1 which is better than any other detector, supervised or unsupervised.

The results of this experiment show that among the supervised detectors, the SMF combined frequency detector has the best performance with the individual frequency confidence maps being added together. In all instances, SMF was able to achieve a TPR of 1 with a relatively low FAR. Although, the results are essentially the same when using either high or low frequency dimensions, low frequency is preferred in one instance. Also, smaller dimensions results in shorter execution time. RX is shown to be the best unsupervised statistics-based

detector using multiplication to create the combined frequency confidence maps with low frequency dimensions. Here, it is able to reach the highest TPR at a FAR of 0.1 before any other detector. Note that the RX and RX_LR results were the same for the all figures and tables in this section.

Table 4-11. Combined Frequency K-Means 4 Original Templates

False Alarm Rate	True Positive Ratio - 16 Targets																												
	κ -Means: 4 Original Templates (Combined Frequency)																												
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		MULT		RX		No Templates												
	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	MULT	ADD											
0.1	0.688	0.625	0.813	0.813	0.5	0.813	0.813	0.75	0.875	0.875	0.125	0.313	0.563	0.438	0.875	0.813	0.125	0.375	0.75	0.938	0.875	0	0	0					
0.3	0.875	0.875	0.875	0.875	0.813	0.875	0.875	1	1	-	0.5	0.438	0.625	0.75	0.75	0.875	0.875	0.813	0.75	0.875	0.938	0.875	0.125	0	0.125				
0.5	0.875	0.875	0.875	0.875	0.875	0.875	0.875	1	1	-	0.688	0.688	0.813	0.813	0.813	0.875	0.875	0.875	0.875	0.813	0.938	0.875	0.313	0.125	0.125				
1	0.938	0.938	0.875	0.875	0.875	0.938	0.938	0.875	0.875	1	1	-	0.875	0.75	0.875	0.875	0.938	0.938	0.938	0.938	0.938	0.938	0.875	0.375	0.125	0.125			
3	1	1	0.875	0.875	1	1	-	1	1	-	0.938	0.938	0.875	0.875	1	1	0.875	0.875	0.938	0.938	0.938	0.938	1	1	0.875	0.5	0.375		
5	1	1	0.875	0.875	1	1	-	1	1	-	0.938	0.938	0.875	0.875	1	1	0.875	0.875	0.938	0.938	0.938	0.938	1	1	0.875	0.563	0.5	0.563	0.438

Table 4-12. Combined Frequency K-Medoids 4 Original Templates

False Alarm Rate	True Positive Ratio - 16 Targets																												
	κ -Medoids: 4 Original Templates (Combined Frequency)																												
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		MULT		RX		No Templates												
	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	MULT	ADD											
0.1	0.625	0.625	0.875	0.875	0.563	0.375	0.875	0.875	0.813	0.813	0.875	0.125	0.25	0.563	0.5	0.875	0.875	0.188	0.313	0.313	0.75	0.938	0.875	0	0	0			
0.3	0.875	0.813	0.875	0.875	0.813	0.75	0.875	0.875	0.813	0.813	0.875	0.625	0.625	0.75	0.688	0.875	0.875	0.75	0.625	0.75	0.875	0.938	0.875	0.125	0	0.125			
0.5	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.875	1	1	-	0.625	0.625	0.813	0.813	0.813	0.875	0.875	0.875	0.875	0.813	0.938	0.875	0.313	0.125	0.125			
1	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.875	1	1	-	0.875	0.813	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.938	0.938	0.875	0.375	0.125			
3	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	1	1	-	0.938	0.938	0.875	0.875	0.938	0.938	0.938	0.938	0.938	0.938	1	1	0.875	0.5	0.375			
5	1	1	0.875	0.875	1	1	-	1	1	-	0.938	0.938	0.875	0.875	1	1	0.875	0.875	0.938	0.938	0.938	0.938	1	1	0.875	0.563	0.5	0.563	0.438

Table 4-13. Combined Frequency K-Means 6 Original Templates

False Alarm Rate	K-Means: 6 Original Templates (Combined Frequency)												True Positive Ratio - 16 Targets												No Templates			
	ACE				ACE ²				SMF				NORMXCORR				CONV				RX				RX_LR			
	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	MULT			
0.1	0.75	0.75	0.875	0.875	0.625	0.563	0.875	0.875	0.813	0.75	-	0.125	0.125	0.188	0.188	0.688	0.688	0.875	0.875	0.125	0.313	0.313	0.75	0.938	0.875	0	0	0
0.3	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.875	-	-	0.938	0.938	-	0.563	0.5	0.625	0.625	0.813	0.813	0.875	0.875	0.938	0.938	0.875	0.125	0	0.125	0.125
0.5	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.875	-	-	0.938	0.938	-	0.688	0.688	0.875	0.875	0.875	0.875	0.875	0.875	0.938	0.938	0.875	0.313	0.125	0.125	0.125
1	0.938	0.938	0.875	0.875	0.938	0.938	-	-	1	1	-	0.875	0.875	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.375	0.375	0.125	0.125
3	0.938	0.938	0.875	0.875	0.938	0.938	-	-	-	-	-	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.5	0.5	0.375	
5	0.938	0.938	0.875	0.875	0.938	0.938	-	-	-	-	-	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.5	0.563	0.438	

Table 4-14. Combined Frequency K-Medoids 6 Original Templates

False Alarm Rate	K-Medoids: 6 Original Templates (Combined Frequency)												True Positive Ratio - 16 Targets												No Templates					
	ACE				ACE ²				SMF				MF				NORMXCORR				CONV				RX				RX_LR	
	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	MULT					
0.1	0.813	0.75	0.875	0.875	0.75	0.625	0.875	0.875	0.813	0.813	-	0.875	0.125	0.313	0.313	0.813	0.813	0.875	0.875	0.125	0.313	0.25	0.75	0.938	0.875	0	0	0		
0.3	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.875	-	-	0.875	0.938	0.938	-	0.625	0.625	0.75	0.875	0.875	0.875	0.813	0.813	0.875	0.938	0.938	0.875	0.125	0	0.125	
0.5	0.938	0.938	0.875	0.875	0.938	0.938	-	-	1	1	-	-	0.688	0.688	0.875	0.875	0.938	0.938	0.875	0.875	0.875	0.875	0.938	0.938	0.875	0.313	0.125	0.125		
1	0.938	0.938	0.875	0.875	0.938	0.938	-	-	1	1	-	-	0.875	0.875	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.375	0.375	0.125		
3	0.938	0.938	0.875	0.875	0.938	0.938	-	-	-	-	-	-	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.5	0.375			
5	0.938	0.938	0.875	0.875	0.938	0.938	-	-	-	-	-	-	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.938	0.875	0.875	0.938	0.563	0.438			

4.4 All Five Sites

The final experiment utilized a much larger number of images and targets to further assert the findings of the results in Sections 4.2 and 4.3 which used a much smaller subset of the entire dataset. Since less than 5% of the entire dataset contain targets, only images with targets are considered in these experiments. Each image contains at least one target. This subset of the data is composed of images taken at five different underwater sites. To perform cross-validation, each site is considered a fold. Table 4-15 shows the number of images in each site along with the number of targets. Site A is not to be confused with Site A* previously mentioned.

Table 4-15. Number of images and targets by site

Site	Number of Images	Number of Targets
A	49	49
B	146	149
C	32	33
D	26	26
E	25	25
Total	278	282

For the template-based detectors, eight K -Means templates were used. K -Means templates are chosen since in a real-world situation the appearance of a target may vary and may even not be known *a priori*. By using K -Means a library of potential targets can be used to estimate the variety.

Section 4.4.1 uses original templates and Section 4.4.2 uses pruned templates. Section 4.4.3 uses combined frequency confidence maps with low frequency dimensions since it provided the best results in previous experiments. The individual high and low frequency confidence maps are combined with addition and multiplication. Additionally, detection is done with and without background thresholding to test the template-based detectors' ability

to distinguish targets from anomalies. Ideally, background thresholding removes anomalous peaks since their confidence values are not greater than the background average. Therefore, the peaks that remain after thresholding and the non-maximum suppression step described in Section 3.4 should then be comprised of majority targets with minimal false alarms. If many false alarms are present, this means the detector is confusing anomalies for targets and this is not desired. In that case, it is a poor performing detector. The unsupervised RX and RX_LR detectors results are also displayed. No background thresholding is performed for these methods since they are unsupervised. Thus, the corresponding ROC curves for these detectors are not included in any of the "with background thresholding" ROC curve figures.

The length of the ROC curve with and without background thresholding is dependent upon the number of peaks that are returned from the non-maximum suppression step. Each time the detector correctly classifies a peak as a target, the TPR increases making the ROC curve more vertical. Incorrect classifications cause the FAR to increase and the curve becomes more horizontal. A ROC curves that is vertical at a FAR of 0 and horizontal at a TPR of 1 is an example of perfect detection. In addition to these ROC curves, tables showing the number of true positives (TP), false alarms (FA), maximum TPR and the TP-to-FA ratio are given. The values in these tables give a better understanding of how well the detectors are performing. By comparing the ROC curves and tables with and without background thresholding we see how well the templates-based and statistics-based detectors are at distinguishing targets from anomalies as opposed to just observing how many are detected. The tables give an overall view of the detectors' performance.

4.4.1 Original Templates

In this section, eight original K -Means templates are used to detect 282 targets. High frequency and low frequency data was used. In the tables that follow, the highest number of TPs and lowest number of FA is desired. The maximum number of TPs is 282 and the number of FAs is infinite. Higher TPR and TP-to-FA ratio values are desired. The TPR is on the interval $[0, 1]$ while the TP-to-FA ratio is on the interval $(-\infty, \infty)$. It should be noted that

the TP-to-FA ratio is a relative value. Here, it is used to compare the detection ability of the individual detectors against each other.

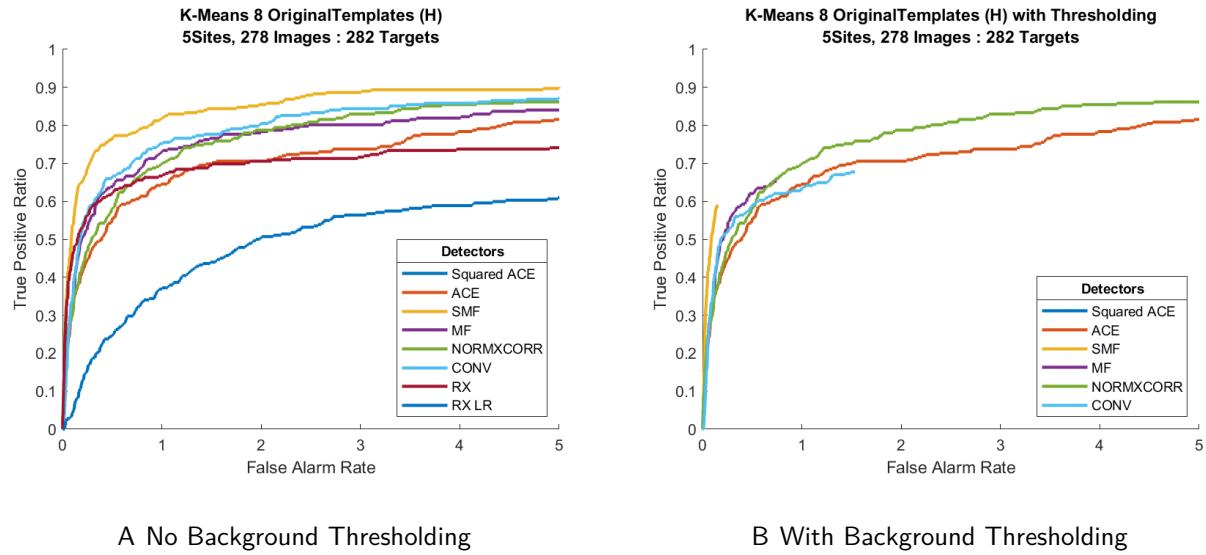


Figure 4-20. High Frequency K -Means 8 Original Templates ROC Curves

Table 4-16. All Five Sites: High Frequency K -Means 8 Original Templates

All Five Sites - 282 Targets														
K -Means: 8 Original Templates (H)														
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		No Templates	
	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	RX	RX_LR
TP	252	237	253	207	269	166	264	184	254	249	267	191	262	217
FA	105,542	25,309	105,870	7,569	105,453	426	93,762	2,076	106,184	20,758	91,366	4,267	105,259	79,766
TPR	0.89	0.84	0.90	0.73	0.95	0.59	0.94	0.65	0.90	0.88	0.95	0.68	0.93	0.77
TP:FA Ratio	0.00	0.01	0.00	0.03	0.00	0.39	0.00	0.09	0.00	0.01	0.00	0.04	0.00	0.00

Figure 4-20A shows the ROC curves of the detectors without using background thresholding. Of the template-based detectors, SMF clearly detects the most targets while ACE detects the least. Table 4-16 shows SMF detecting 269 targets reaching a TPR of 0.95. However, this detector also recognizes 105,453 anomalies and has a TP-to-FA ratio of zero. Using no background thresholding, many of the other detectors perform well reaching a TPR of at least 0.89 though all have a TP-to-FA ratio of zero. Of the unsupervised statistics-based

detectors, RX does the best detecting 262 targets while RX_LR detects only 217. Both detectors also have TP-to-FA ratios of zero meaning they also mistake many anomalies as targets.

Figure 4-20B shows the ROC curves of the detectors with background thresholding. The number of false alarms are significantly reduced. Although the TPRs are lower, the TP-to-FA ratios increase with SMF reaching the highest at 0.39 with the least number of FAs of all the detectors when using thresholding. The ROC curves for SMF in this case is unusual since it stops at a TPR of 0.59. Other detectors stop at low TPRs as well. This is due to the number of peaks returned after non-maximum suppression and thresholding.

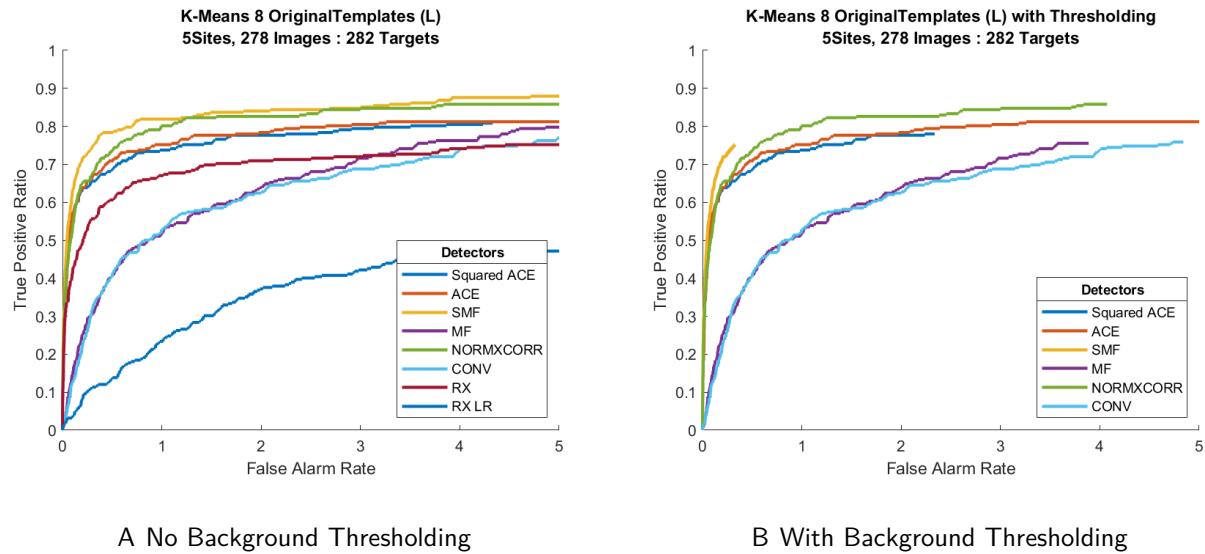


Figure 4-21. Low Frequency *K*-Means 8 Original Templates ROC Curves

Table 4-17. All Five Sites: Low Frequency *K*-Means 8 Original Templates

	All Five Sites - 282 Targets												No Templates	
	K-Means: 8 Original Templates (L)												RX	RX_LR
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV			
	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh		
TP	242	240	240	220	257	212	254	213	253	242	248	214	258	173
FA	53,664	42,008	54,465	6,496	53,988	897	41,525	10,806	53,196	11,319	40,854	13,456	53,420	42,436
TPR	0.86	0.85	0.85	0.78	0.91	0.75	0.90	0.76	0.90	0.86	0.88	0.76	0.91	0.61
TP:FA Ratio	0.00	0.01	0.00	0.03	0.00	0.24	0.01	0.02	0.00	0.02	0.01	0.02	0.00	0.00

Next, the same experiment was done using the low frequency templates and SAS images. As seen in previous experiments, the MF and CONV detectors suffer when using low frequency data. This is observed in the ROC curves of Figure 4-21. In Figure 4-21A, SMF again has the most success detecting targets with a TPR of 0.91 with no background thresholding. Again, SMF and the other template-based detectors have a TP-to-FA ratio of zero which indicates that they are detecting many anomalies as well. The curves in this figure are vertical near the lowest FAR which is desirable of the detectors. Near the TPR of 0.85 the detectors begin to become more horizontal. This is due to the false alarms since more are being detected as targets at that TPR. The RX and RX_LR detector's performance also declines when using low frequency data. RX detects 258 target and RX_LR detects 173. They achieve TPRs of 0.91 and 0.61, respectively though both has TP-to-FA ratios of zero.

From these results it shows that with background thresholding the TPRs for all of the detectors decrease as well as the number of false alarms. This trend is seen in both high and low frequency tests. For SMF, this reduction in the number of false alarms is the most significant going from 53,988 to 897. The ratio of TP-to-FA also increases for all detectors with SMF having the most significant increase from 0 to 0.24 with thresholding meaning that it is able to detect the most targets with the least number of false alarms. Another key point is that when no thresholding is used, the number of false alarms is nearly cut in half when using low frequency as opposed to high frequency. This means that although less targets are detected with low frequency, the number of anomalies detected as targets are also significantly reduced.

4.4.2 Pruned Templates

The same experiment was done again using eight K-Means pruned templates. The results are shown in the following Figures and Tables. Figure 4-22 shows the high frequency data ROC curves and Figure 4-23 shows the low frequency data ROC curves. The RX and RX_LR results shown in this section are the same as those described in the previous section since they are unsupervised statistics-based detectors.

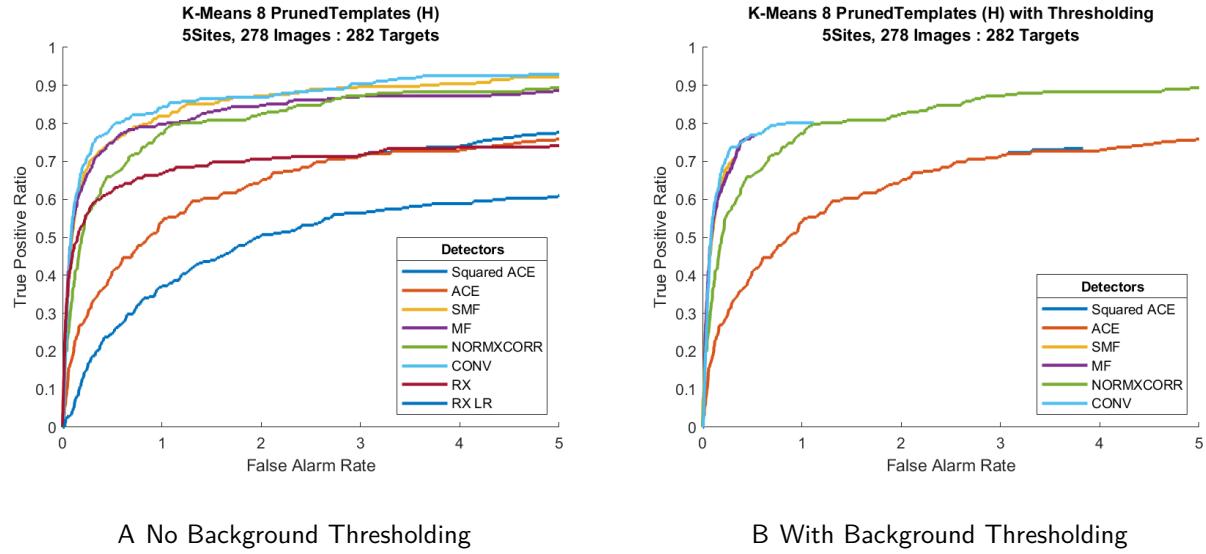


Figure 4-22. High Frequency K -Means 8 Pruned Templates ROC Curves

Table 4-18. All Five Sites: High Frequency K -Means 8 Pruned Templates

	All Five Sites - 282 Targets												No Templates	
	K-Means: 8 Pruned Templates (H)												No Templates	
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		RX	RX_LR
	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh		
TP	241	237	244	207	272	200	270	216	261	258	272	226	262	217
FA	106,987	31,781	107,358	10,651	107,727	921	97,495	1,461	106,396	24,836	97,399	3,113	105,259	79,766
TPR	0.85	0.84	0.87	0.73	0.96	0.71	0.96	0.77	0.93	0.91	0.96	0.80	0.93	0.77
TP:FA Ratio	0.00	0.01	0.00	0.02	0.00	0.22	0.00	0.15	0.00	0.01	0.00	0.07	0.00	0.00

In Figures 4-22A and 4-22B, ACE and ACE^2 again show degraded performance when using pruned templates. With and without background thresholding their performances are nearly identical according to the ROC curves. The remaining detectors all have similar performances. However, by examining Table 4-18, some of the high frequency differences become more apparent. When background thresholding is not used, SMF and CONV are able to detect 272 targets at a TPR of 0.96. SMF typically outperforms the other detectors but here it has 107,727 false alarms whereas the CONV detector only has 97,399. This means that CONV is slightly better than SMF at distinguishing targets from anomalous objects.

When background thresholding is used with high frequency data there is again a significant reduction in the number of false alarms and targets detected. SMF experiences the largest reduction and achieves the highest TP-to-FA ratio of 0.22 though it does not detect the most targets. SMF is only able to detect 200 targets but with 921 false alarms. NORMXCORR however has the highest TPR of all the detectors at 0.91, finding 258 targets but with 24,836 to get a TP-to-FA ratio of 0.01. This ratio is lower than the ratio of SMF only due to the higher number of false alarms. The lowest TPR among all of the detectors using background thresholding is 0.71.

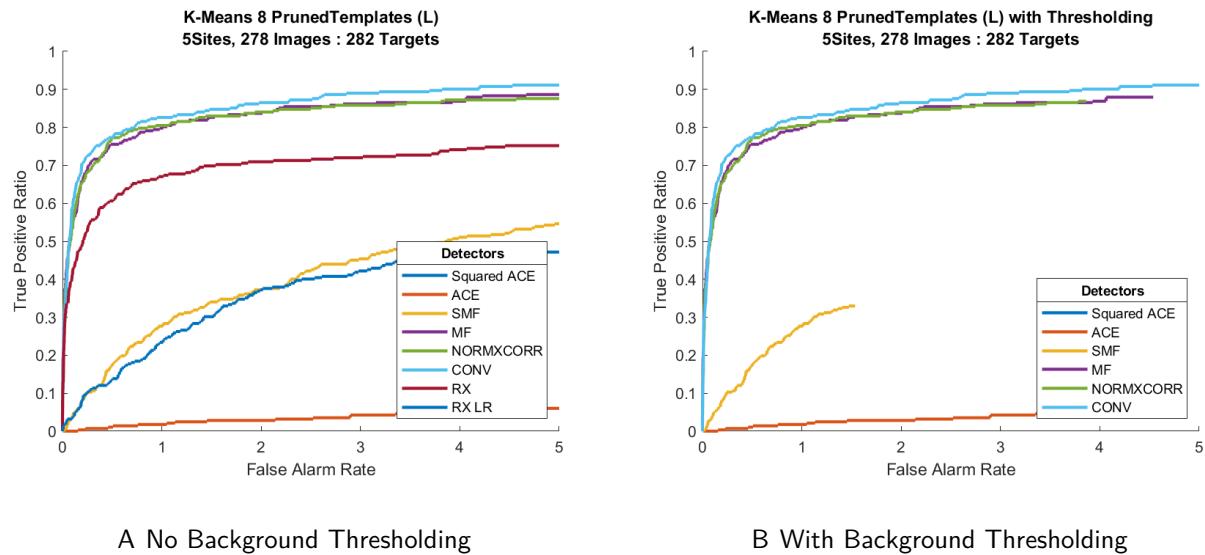


Figure 4-23. Low Frequency *K*-Means 8 Pruned Templates ROC Curves

Table 4-19. All Five Sites: Low Frequency *K*-Means 8 Pruned Templates

	All Five Sites - 282 Targets												No Templates	
	K-Means: 8 Pruned Templates (L)						CONV						RX	RX_LR
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV			
	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh		
TP	133	16	137	9	200	93	267	248	259	245	267	259	258	173
FA	56,050	13,324	55,993	6,728	55,850	4,270	49,683	12,613	53,100	10,736	49,300	17,075	53,420	42,436
TPR	0.47	0.06	0.49	0.03	0.71	0.33	0.95	0.88	0.92	0.87	0.95	0.92	0.91	0.61
TP:FA Ratio	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.02	0.00	0.02	0.01	0.02	0.00	0.00

Figure 4-23 shows the ROC curves of the experiment using eight K -Means pruned templates on low frequency data with and without background thresholding. Compared to the other detectors, ACE and ACE^2 performed the worst reaching a TPR of only 0.47 and 0.49, respectively without thresholding. The average number of FAs for these detectors is 56,021, thus their TP-to-FA ratio is zero. The MF and CONV detectors found the most targets reaching a TPR of 0.95, showing improvement again when pruned templates are used. The average number of FAs for these detectors is 49,491, thus their TP-to-FA ratio is 0.01. SMF did not have the best performance in this instance, achieving a TPR of only 0.71 with 55,850 FAs and a TP-to-FA ratio of zero.

When background thresholding is used, the TPRs for all of the detectors decrease, some much more than others. ACE and ACE^2 's TPRs decrease to an average of 0.05 with the average number of FAs being 10,026. The CONV detector had the best performance with a TPR of 0.92 and 17,075 FAs. MF and NORMXCORR have an average TPR of 0.88 with an average of 11,675 FAs. SMF had a TPR of 0.33 with 4,270 FAs. It was the worst performing detector in this case. The TP-to-FA ratio for the SMF, MF, NORMXCORR and CONV detectors had a slight increase to 0.02 compared to the results without background thresholding. This is an instance where the TP-to-FA ratio indicates the relativity of the detector's performance. The ratio essentially is a flag for a second look by a human operator as opposed to a measure of how well a detector is performing. The trends of this experiment mirrors that of the experiments in Subsection 4.2.2.

4.4.3 Combined Frequency

Finally, combined frequency confidence maps were used on the five site dataset. The individual frequency confidence maps were combined using addition and multiplication. Low frequency dimensions were used for experimentation. Again, a comparison on thresholding is done on the supervised template-based detectors.

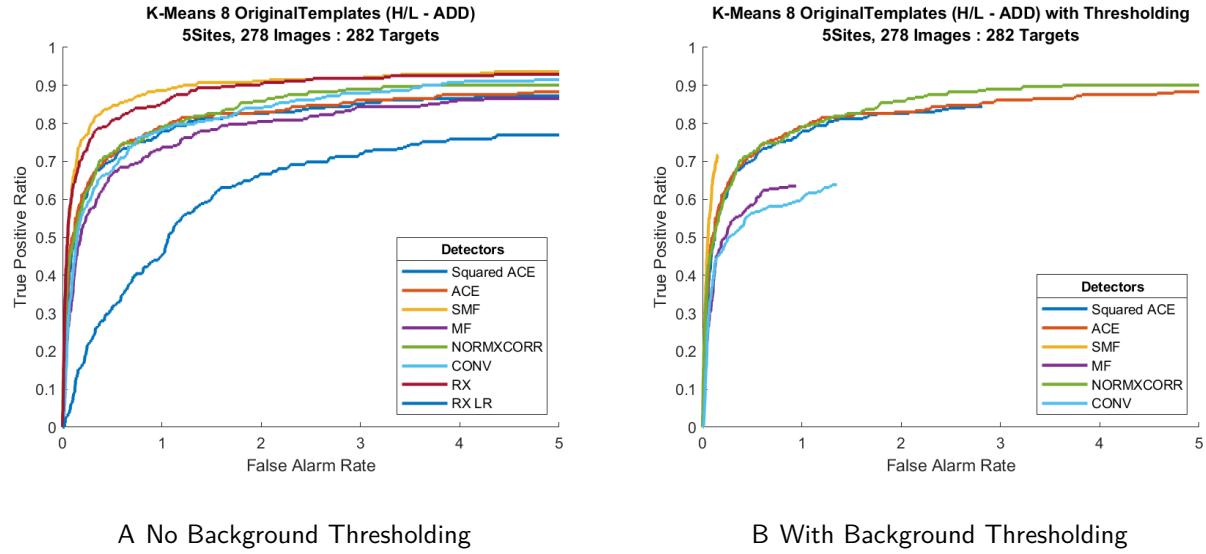


Figure 4-24. *K*-Means Addition Combined Frequency 8 Original Templates ROC Curves

Table 4-20. All Five Sites: Combined Frequency *K*-Means 8 Original Templates using Addition

All Five Sites - 282 Targets														
K-Means: 8 Original Templates (H/L) - Addition														
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		No Templates	
	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	RX	RX_LR
TP	256	252	255	238	272	202	261	179	259	255	268	180	265	242
FA	54,325	24,252	54,564	7,832	54,433	427	45,653	2,622	54,237	15,427	45,432	3,762	53,895	49,515
TPR	0.91	0.89	0.90	0.84	0.96	0.72	0.93	0.63	0.92	0.90	0.95	0.64	0.94	0.86
TP:FA Ratio	0.00	0.01	0.00	0.03	0.00	0.47	0.01	0.07	0.00	0.02	0.01	0.05	0.00	0.00

Figure 4-24 shows the ROC curves of the experiment when addition is employed to combine the individual frequency confidence maps with and without background thresholding. The ROC curve in Figure 4-24A shows SMF as the standout detector with the remaining detectors performing similarly with no background thresholding. If attention is directed to Table 4-20, it shows that SMF indeed has the highest TPR of 0.96 with 54,433 FAs. The lowest TPR among the supervised detectors without background thresholding is 0.92. The only detectors to have a TP-to-FA ratio above zero are MF and CONV which averaged 45,543 FAs. The unsupervised RX and RX_LR detectors reach a TPR of 0.94 and 0.86, respectively, with both having TP-to-FA ratios of zero. Again, SMF proves to be the best supervised detector

and RX is the best unsupervised detector when using addition to combine the high and low frequency confidence maps.

When background thresholding is used, three of the detectors suffer the most in regards to their TPR. As shown in Figure 4-24B, the SMF, MF and CONV ROC curves cut off. The lowest TPR of these detectors is 0.63 and they average 2,270 FAs. Impressively, SMF only has 427 FAs though only a TPR of 0.72. It also has the highest TP-to-FA ratio of 0.47.

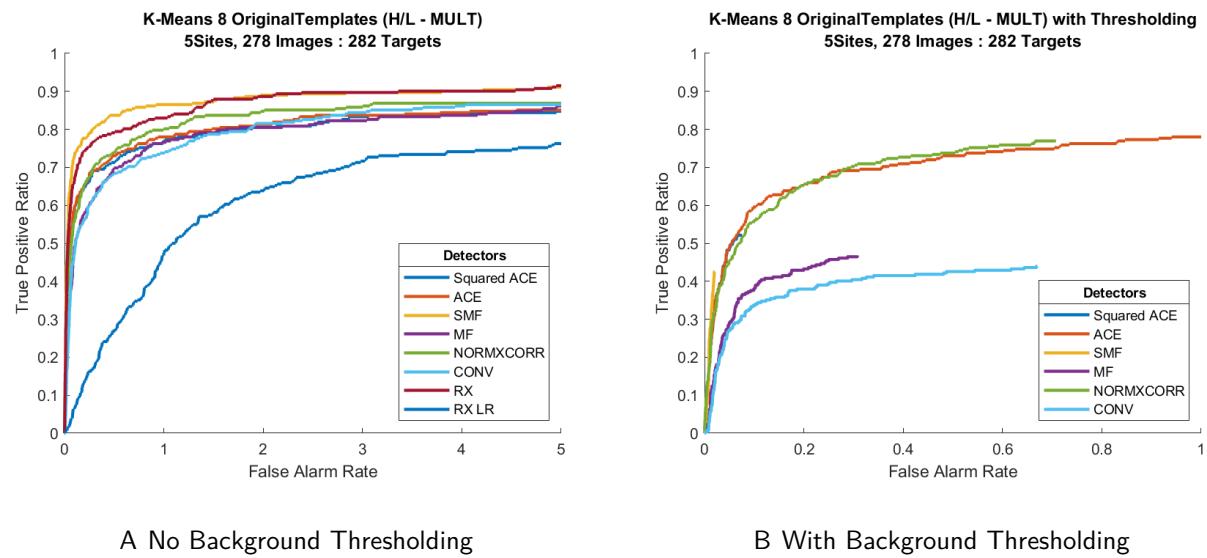


Figure 4-25. *K*-Means Multiplication Combined Frequency 8 Original Templates ROC Curves

Table 4-21. All Five Sites: Combined Frequency *K*-Means 8 Original Templates using Multiplication

All Five Sites - 282 Targets															
K-Means: 8 Original Templates (H/L) - Multiplication														No Templates	
	ACE		ACE ²		SMF		MF		NORMXCORR		CONV		RX	RX_LR	
	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh	No Thresh	Thresh			
TP	250	220	251	147	266	120	261	132	250	217	260	124	261	229	
FA	53,166	3,282	53,568	210	53,551	56	45,705	855	53,331	1,967	44,408	1,862	52,914	36,148	
TPR	0.89	0.78	0.89	0.52	0.94	0.43	0.93	0.47	0.89	0.77	0.92	0.44	0.93	0.81	
TP:FA Ratio	0.00	0.07	0.00	0.70	0.00	2.14	0.01	0.15	0.00	0.11	0.01	0.07	0.00	0.01	

The final comparison results are shown in Figure 4-25. The ROC curves are from the experiment using combined frequency confidence maps that were combined using multiplication. The TPRs for this experiment are lower than those for when addition is used to combine the individual frequency confidence maps. In Figure 4-25A again SMF outperforms the other template-based detectors reaching a TPR of 0.94 with 53,551 FAs and a TP-to-FA ratio of zero. The detectors with the lowest TPR is ACE, ACE² and NORMXCORR which have a TPR of 0.89 and TP-to-FA ratio of zero. The ROC curves for these detectors are very similar. CONV and MF have TP-to-FA ratios Of 0.01 and TPRs of 0.92 and 0.93, respectively. Along with SMF these supervised detectors did the best at finding the most targets. RX and RX_LR have comparable results reaching TPRs of 0.93 and 0.81, respectively. RX_LR in this instances has a TP-to-FA ratio of 0.1, higher than RX meaning it may be better at distinguishing here.

The ROC curves generated by using background thresholding are shown in Figure 4-25B. The highest TP-to-FA ratio of 2.14 using combined frequency confidence maps occurs in this experiment for SMF. Again, the ratio value serves as a flag to further investigate the performance of the detector. The value is high in this case because only 56 FAs are found. The TPR for the detector however, is 0.43 which can be considered unsatisfactory. The ACE detector achieves the highest TPR of 0.78 with only 3,282 FAs and a TP-to-FA ratio of 0.07 while ACE² has a TPR of 0.52 and a ratio of 0.70. NORMXCORR reaches a TPR of 0.77 with a TP-to-FA ratio of 0.11. MF and CONV reach TPRs of 0.47 and 0.44, respectively, with ratios of 0.15 and 0.07.

CHAPTER 5 CONCLUSION

In this thesis, three different experiments were conducted to study anomaly and target detection on an underwater SAS imagery dataset. Supervised template matching and unsupervised statistical methods were compared to assess their ability to detect targets within the pre-labeled dataset with the goal of producing a low number of false alarms or anomalies with a high true positive ratio. N -fold cross validation was used in all experiments to prevent contamination of the testing data with training data. ROC curves and TPR tables were used to display the results of each experiment. A table of TP and FA metrics accompany the results of the third experiment.

The first two experiments used a small subset of the dataset, called Site A*, to investigate the effects of the number of templates, type of templates, frequency of the SAS images and the method used to produce final confidence maps. Additionally, K -Means and K -Medoids templates were compared. The second experiment used the Site A* dataset and investigated the benefits of combining the individual high and low frequency confidence maps for detection. K -Means and K -Medoids templates were again used for comparison. The third and final experiment used a larger subset of the data which included imagery from five different underwater sites labeled as Site A-E. Only K -Means templates were used in this experiment. The type of template used was varied and the results are compared with and without the use of background thresholding to investigate the detectors' ability to distinguish targets from anomalies. In all experiments the supervised template-based detectors are compared against the unsupervised RX and RX_LR detectors that required minimal parameter changes.

Section 4.2 details the results of the first experiment where a comparison of K -Means versus K -Medoids estimated templates were used to detect targets. Neither K -Means nor K -Medoids showed any significant superiority as a template estimating method. These results are shown together with the unsupervised detection results. In Sections 4.2.1 and 4.2.2, the effect of original and pruned target templates were studied. By comparing the ROC curves

and TPR tables given in each section, it is best to use original templates when attempting to perform detection on this SAS dataset. Four and six original and pruned target templates were used and various parameters were changed to study their effects on the detection performance of each detector. The number of templates used was varied to see if four or six templates were able to create more distinct target templates capable of capturing the variability of the targets. Overall, four templates estimates the best results whether they are original or pruned target templates. Most detectors reached its highest TPR when high frequency templates and images were used. For this reason, high frequency data is most advantageous for performing target detection than low frequency. However, with low frequency data the detectors seemed to be more precise at lower FARs since the TPR values were generally higher than those for high frequency data.

Section 4.3 details the results of the second experiment where combined frequency confidence maps were employed to explore the benefits of combining the individual high and low frequency confidence maps for any SAS image. Again, a comparison between *K*-Means and *K*-Medoids was done in addition to varying the number of templates. Section 4.3.1 shows the results from using four original templates and Section 4.3.2 shows the results from using six original templates. In order to combine the individual high and low frequency confidence maps, which have differing dimensions, one map could either be stretched and the other shrunk or vice versa. Then, the maps are combined either with addition or multiplication. By varying the final combined frequency confidence maps' dimensions and method for combination, each SAS image ends up with four different combined frequency confidence maps for detection.

According to the results of this experiment, the dimensions of the combined frequency confidence maps matters when a preference in desired FAR is specified. For most of the experiments, the TPR values for the combined frequency confidence maps with high frequency dimensions versus low frequency dimensions are essentially identical. This means that stretching or shrinking one versus the other individual frequency confidence maps has no affect on the outcome of the final combined frequency confidence map. However, the method

for combining the map produces very different results. When the maps are combined using addition, the TPRs were higher for all detectors. A majority of the detectors are able to detect all of the targets, whereas when multiplication is used only 14 targets were detected for all of the detectors.

For this reason, multiplication may be the best method for detection with combined frequency confidence maps when a low FAR is desired since at the lowest FARs, for both dimensions, the TPRs are better than those using addition. Four proved to be the best number of templates to use to create the combined frequency confidence maps. Using both *K*-Means and *K*-Medoids, all of the detectors are able to detect all 16 targets when the maps are combined using addition. When six templates are used with addition, only SMF is able to do so for both template methods. Additionally, missing TPR values occur in the TPR tables of the combined frequency confidence maps results. This indicates that there were no more peaks returned from the non-maximum suppression step. The peaks account for some combination of targets and false alarms or anomalies. This means that no more anomalies were detected but also no more targets.

Section 4.4 details the third experiment which combines findings from the previous two experiments and investigates the results of using a larger dataset. The first two experiments were concerned with observing the effects of the different parameters for creating the confidence maps used to detect the targets. This experiment puts this knowledge together to see just how well the detectors are at actually distinguishing the targets from anomalies with the confidence maps. The quality of detection is explored. The results of this experiment can be considered more reliable because of the larger dataset and varied seafloor characteristics of sites A-E. There are also 282 targets present within this dataset as opposed to only 16 in the Site A* dataset. This enables the results to be a lot clearer and more consistent though they do share similarities to the experiments of Sections 4.2 and 4.3.

The question this experiment asks is how well are the templates capturing the image of the targets and producing reliable confidence maps in order to detect the targets in a

supervised manner. To make the answer to this clearer, detection is done without and with background thresholding in an attempt to further reduce the number of false alarms or anomalies detected. In an unsupervised manner, the experiment asks how well are the statistics-based detectors able to distinguish the surrounding background in an image from an object present within it. Eight original and pruned *K*-Means templates were used with high and low frequency confidence maps along with combined frequency confidence maps to perform detection. ROC curves were again used to capture the TPR vs FAR trends of each of the detectors but instead of a TPR table of values and corresponding FAR values, TP and FA metrics are used to describe performance.

The ROC curves exhibit similar trends as those in the previous experiments in regards to individual detector performance, template type performance (original vs pruned), and frequency performance. In the figures, there are many situations where the curves appear vertical but cuts off before the TPR can reach 1. This happens when background thresholding is used. It means that background thresholding removed many of the peaks but the detector wasn't able to find all of the targets. Results like this can be attributed to the quality of the templates and detector. However, if the ROC curve cuts off but appears horizontal, this means that the detector again didn't find all of the targets but also did poorly at distinguishing them from anomalies which produces false alarms. The detectors are giving targets and anomalies high confidence values. These objects that aren't targets are basically considered very similar to them which brings into question the quality of the templates and detector. However, if the objective is to detect anomalies as opposed to targets, this would be a good example of anomaly detection. In practice, these anomalies could then be used in the next stage of target detection since they appear similar.

Without thresholding the number of FAs are generally doubled in addition to the detectors being able to find a majority of the targets. About 80% of the targets are detected with FARs less than 0.5 while TPRs higher than 80% show increased FARs. With these findings, thresholding can be thought of as target detection and no thresholding as anomaly detection.

There is a clear trade-off between the number of targets detected and the number of false alarms detected with and without the use of thresholding. The TP and FA metric tables put the number of targets detected and anomalies found into perspective. In addition to the counts of TPs and FAs, the TPR and TP-to-FA ratio is given.

A desirable detection would have a TP near 282, FA near 0, TPR near 1 and TP-to-FA ratio as high as possible. Although, the TPR may reach a value of 1, it does not mean that the detector is good at distinguishing targets from anomalies. For instance, the number of FAs could be very high meaning that the detector labels many things as anomalies. The TP-to-FA ratio aids in identifying this problem and serves as a flag to investigate further the performance of the detector. There are three options for the value of this ratio. When the ratio is greater than 0, there is a relatively small amount of FAs compared to the number of TPs. However, the number of TPs could also be small too. When the number of FAs is much lesser than the number of TPs, the ratio increases. This is a situation where a closer look is needed. It can also be undefined meaning that there were no FAs or anomalies. Finally, the ratio could be equal to or very close to zero meaning that there are a large number of FAs compared to the number of TPs.

In conclusion, template-based matching and statistical detection methods were studied with various parameters to investigate supervised and unsupervised anomaly and target detection. Combined frequency data was determined to be the most advantageous source for detecting targets with low frequency dimensions showing the most promising results when precision is necessary. The use of four original templates are preferred over pruned templates with neither *K*-Means nor *K*-Medoids showing strong dominance in performance. Significant target distinguishing capabilities however, are not clear in the results.

The supervised SMF detector and unsupervised RX detector consistently produced the best detection results. For this reason, if time is not an issue, the SMF template-based detector with 4 original templates and combined frequency confidence maps using low frequency dimensions would be the best system for detection. The results in this thesis showed

that SMF is a versatile and reliable supervised detector; always among the top detectors regardless of parameter setting. However, if time and storage are important factors in the design of a system, the unsupervised RX detector would be best since no training is necessary, thus it runs much faster. A significantly less number of confidence maps is required for detection which reduces storage and computation time.

Detecting FAs or anomalies can be a good and bad thing however, depending on the objective. Supervised and unsupervised systems can be designed to favor one more than the other with the trade-off being certainty versus precaution. A system can be designed to return objects that are determined to be targets with high confidence or return all anomalies that raise concern. In practice, operators would consider it a pain, but it allows for extra care to be taken to ensure that a target isn't potentially missed.

REFERENCES

- [1] P. Lyons, D. Suen, A. Galusha, A. Zare, and J. Keller, "Comparison of prescreening algorithms for target detection in synthetic aperture sonar imagery," International Society for Optics and Photonics, pp. 387 – 394, April 2018. [Online]. Available: <https://doi.org/10.1117/12.2305175>
- [2] H. Kwon and N. M. Nasrabadi, "Kernel rx-algorithm: a nonlinear anomaly detector for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, Feb 2005. [Online]. Available: <https://ieeexplore.ieee.org/document/1386510/>
- [3] V. Saligrama, E. Arias-Castro, R. Chellappa, A. O. Hero, R. Nowak, and V. V. Veeravalli, "Introduction to the issue on anomalous pattern discovery for spatial, temporal, networked, and high-dimensional signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 1–3, Feb 2013.
- [4] D. Huang, D. Mu, L. Yang, and X. Cai, "Codetect: Financial fraud detection with anomaly feature detection," *IEEE Access*, vol. 6, pp. 19 161–19 174, 2018.
- [5] M. Hauskrecht, I. Batal, M. Valko, S. Visweswaran, G. F. Cooper, and G. Clermont, "Outlier detection for patient monitoring and alerting," *Journal of biomedical informatics*, vol. 46, no. 1, pp. 47–55, feb 2013. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/22944172><https://www.ncbi.nlm.nih.gov/pmc/PMC3567774/>
- [6] L. Li, "Anomaly detection in airline routine operations using flight data recorder data," Ph.D. dissertation, Massachusetts Institute of Technology, 2013. [Online]. Available: <http://hdl.handle.net/1721.1/82498>
- [7] S. S. Kim and A. L. Narasimha Reddy, "Image-based anomaly detection technique: Algorithm, implementation and effectiveness," *Selected Areas in Communications, IEEE Journal on*, vol. 24, pp. 1942 – 1954, 11 2006.
- [8] N. Vashistha, M. T. Rahman, H. Shen, D. L. Woodard, N. Asadizanjani, and M. Tehranipoor, "Detecting hardware trojans inserted by untrusted foundry using physical inspection and advanced image processing," *Journal of Hardware and Systems Security*, vol. 2, no. 4, pp. 333–344, Dec 2018. [Online]. Available: <https://doi.org/10.1007/s41635-018-0055-0>
- [9] A. Galusha, G. Galusha, J. M. Keller, and A. Zare, "A fast target detection algorithm for underwater synthetic aperture sonar imagery," pp. 10 628 – 10 628 – 13, 2018. [Online]. Available: <https://doi.org/10.1117/12.2304976>
- [10] I. S. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1760–1770, Oct 1990.

- [11] "How fast does sound travel?" Apr 2018. [Online]. Available: <https://dosits.org/science/movement/how-fast-does-sound-travel/>
- [12] (2018) What is sonar? [Online]. Available: <https://oceanservice.noaa.gov/facts/sonar.html>
- [13] (2018) Types of sonar. [Online]. Available: <http://ww2.simrad-yachting.com/en-US/Products/Echosounders/Types-of-Sonar/>
- [14] D. B. Fissel, R. A. J. Chave, M. Clarke, P. Johnston, K. Borg, J. R. Marko, E. Ross, J. Buermans, and M. Stone, "Advances in moored upward looking sonar systems for long term measurement of arctic ice and oceanography," in *2013 OCEANS - San Diego*, Sept 2013, pp. 1–7. [Online]. Available: <https://ieeexplore.ieee.org/document/6741279/>
- [15] (2018) Side scan sonar. [Online]. Available: https://www.usna.edu/Users/oceano/pguth/md_help/geology_course/side_scan_sonar.htm
- [16] M. P. Hayes and P. T. Gough, "Synthetic Aperture Sonar: A Review of Current Status," *IEEE Journal of Oceanic Engineering*, vol. 34, no. 3, pp. 207–224, jul 2009. [Online]. Available: <http://ieeexplore.ieee.org/document/5191242/>
- [17] R. E. Hansen, "Introduction to synthetic aperture sonar," in *Sonar Systems*, N. Z. Kolev, Ed. Rijeka: IntechOpen, 2011, ch. 1. [Online]. Available: <https://doi.org/10.5772/23122>
- [18] J. W. Kaeli, "Real-time anomaly detection in side-scan sonar imagery for adaptive auv missions," in *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*, Nov 2016, pp. 85–89. [Online]. Available: <https://ieeexplore.ieee.org/document/7778653/>
- [19] D. P. Williams, "Image-quality prediction of synthetic aperture sonar imagery," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 2114–2117.
- [20] D. Williams, "Fast target detection in synthetic aperture sonar imagery: A new algorithm and large-scale performance analysis," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 1, pp. 71–92, Jan 2015.
- [21] F. C. Crow, "Summed-area tables for texture mapping," in *ACM SIGGRAPH computer graphics*, vol. 18, no. 3. ACM, 1984, pp. 207–212.
- [22] X. Du, A. Seethapalli, H. Sun, A. Zare, and J. T. Cobb, "Environmentally-adaptive target recognition for sas imagery," pp. 10182 – 10182 – 15, 2017. [Online]. Available: <https://doi.org/10.1117/12.2262688>
- [23] G. C. Stefania Matteoli, Marco Diani, "Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images," *Optical Engineering*, vol. 49, pp. 49 – 49 – 16, 2010. [Online]. Available: <https://doi.org/10.1117/1.3386069>
- [24] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157 vol.2. [Online]. Available: <https://ieeexplore.ieee.org/document/790410/>

- [25] A. Zare, C. Jiao, and T. Glenn, "Discriminative multiple instance hyperspectral target characterization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 10, pp. 2342–2354, Oct 2018.
- [26] K. Turkowski, "Turkowski filters for common resampling tasks 10 april 1990 filters for common resampling tasks."

BIOGRAPHICAL SKETCH

Princess Lyons completed her dual Bachelor of Science in Electrical and Computer Engineering as a Honors Scholar with minors in Mathematics, Computer Science and Spanish from the University of Missouri in Columbia. She excelled academically, earning a Boeing Scholarship and maintaining the University of Missouri's Diversity Award. Princess also spent two summer internships as a Software Engineer for Lockheed Martin in King of Prussia, PA. Her community, social and academic endeavors led to her being selected as one of the first honorees for the Celebration of Women in Engineering Banquet at the University of Missouri.

After completing her undergraduate studies, Princess received her Master of Science in Electrical and Computer Engineering at the University of Florida in 2019. There she studied machine learning and image processing under the supervision of Dr. Alina Zare. Her research focused on underwater anomaly and target detection in synthetic aperture sonar imagery.