FINAL REPORT

Belief Theoretic Multi-Sensory Data Fusion for Underwater UXO Identification

SERDP Project MR-1660

FEBRUARY 2011

Shahriar Negahdaripour
K. Premartne
D.T.L. Wickramarathne
University of Miami

P.-P. Beaujean
L. N. Brisson
Florida Atlantic University

This document has been cleared for public release

SERDP
DOD • EPA • DOE

This report was prepared under contract to the Department of Defense Strategic Environmental Research and Development Program (SERDP). The publication of this report does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official policy or position of the Department of Defense. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Department of Defense.

CONTENTS

Cover Page

Contents

List of Figures i

List of Tables iii

List of Acronyms and Abbreviations iv

List of Symbols v

Executive Sumary 1

I Introduction 2

II Project Overview 2
- II-A Background . . . 2
- II-B Objectives . . . 3
- II-C Proposed Plan and Testing . . . 4
- II-D Revised Plan . . . 5

III Overview of Technical Work and Accomplishments 6

IV Project Tasks 8
- IV-A Data Acquisition . . . 8
  - IV-A1 Tank Data . . . 9
  - IV-A2 SERDP/ESTCP Data . . . 11
  - IV-A3 Marina Data . . . 13
- IV-B Data Processing . . . 13
  - IV-B1 Test Site Mosaic Generation . . . 13
  - IV-B2 3-D Reconstruction . . . 14
  - IV-B3 Optical and Acoustic Image Registration and 3-D Reconstruction 14
  - IV-B4 MAFR-TM Technique . . . 19
- IV-C Multi-Modal DS-Based Data Fusion . . . 21

V Dempster-Shafer (DS) Theory Related Work 22
- V-A Basic Notions . . . 22
- V-B Evidence Combination . . . 23
- V-C Conditional Update Equation (CUE) . . . 23
- V-D DS Theoretic Characterization of Optical and Acoustic Data and Problem Formation . . . 24

VI Results 25
VI-A Test Site Mosaic Generation . . . 25
VI-B 3-D Reconstruction from Stereo Data . . . 27
VI-C Registration and 3-D Reconstruction of Occluding Contours . . . 28
VI-D MAFR-TM Technique . . . 29
VI-D1 Target Detection . . . 35
VI-D2 Classification . . . 36
VI-D3 Target Orientation . . . 39
VI-D4 Target Location . . . 41
VI-E DS-based Fusion Technique . . . 42
VI-E1 Image Preprocessing and Shadow Marking . . . 43
VI-E2 Evidence Modeling . . . 43
VI-E3 Combining Evidence Sources . . . 46

VII Recommendations and Future Work 49
VII-A Optical Imaging . . . 49
VII-B FS Sonar Imaging . . . 50
VII-C DS Theoretic Evidence Fusion . . . 51
VII-C1 Shadow Analysis . . . 51
VII-C2 Evidence Filtering . . . 51
VII-C3 Additional Sensor Modalities and Complementary Classification Techniques . . . 51
VII-C4 Reliability Issues . . . 51
VII-D Other Issues . . . 52

VIII Publications 52

References 53

Appendix A; Details of MAFR-TM Algortihm 55
A.1 Overview . . . 55
A.2 Source Image and Templates . . . 56
A.3 Correlation . . . 57
A.4 Detection and Initial Localization . . . 58
A.5 Classification and Orientation . . . 59
A.6 Final Localization . . . 59
A.7 MAFR-TM Optimization . . . 60

Appendix B: Sample Sonar Images of Various Targets in the Pool 61

Appendix C: Sample Images from Our Data Set 62

LIST OF FIGURES

1 FAU's outdoor tank facility in Boca Raton (left) and marina at Sea Tech Campus (right). 5

2 Camera assemblies for (a) pool and (b) marina deployment. 7

3 Various targets with different shapes and reflectance properties; cinder block, dumbbell, as well as steel, Aluminum and PVC pipes 8

4 Three cylindrical targets used in the construction of the source image data base for the testing of the MAFR-TM technique 9

5 Tank (top) and marina (bottom) setups. 10

6 Sample data of steel cylinder acquired in outdoor pool with (a) DIDSON and (b) BlueView, and (c) with an optical camera. 11

7 Setup for target imaging with clutter in both outdoor tank and marina. 11

8 Sample images acquired in FAU marina, with (a,b) BlueView at 120 and 160 [deg] and (c,d) DIDSON at 0 and 120 [deg] relative to the target. (e,f) optical stereo pair at about 2 [ft] distance. 12

9 (a,b) Overlapping optical and sonar views; (c) registered over the plane defined by the two PVC pipes. (column d) selected 3-D targets with different visual characteristics in the optical image and (column e) registered views. In absence of a priori information about number and visual properties of these targets, segmentation results is rather poor by applying an optimum global threshold over the entire image based on simplistic bimodal distribution model (column f)). Despite coarse image alignment of these 3-D targets, locating each in either/sonar image modality enhances ability to locate it in other/optical view, by adjusting detection parameter(s) optimally based on intensity distribution within region around each target (column g). 16

10 (a-b) One sample sonar and optical images from a data set acquired in UM indoor tank with clear water. (c) enhanced optical image for better visualization. (d-g) A particular object (labeled 7), and segmentation results by apply k-means with 2-4 clusters. Bimodal distribution for 2 clusters can produce very good results when applied to a small region around the potential target. Region can be over segmented when a larger number of clusters is assumed. 18

11 Regions around each object taken from different sonar images in the data set, showing that an object's shadow provides cues about object's shape and its occluding contour. 19

12 Reconstruction of occluding contour from shadow boundary of sonar image C, by tracing from boundary points P along iso-azimuth frontal acoustic beams a to points P' on the occlusion boundary O. 20
13 Sample raw images used for construction of site mosaic and computer camera trajectory. 26
14 Camera trajectory for sample mosaic of natural terrain constructed from data in Andros Island, Bahamas. 26
15 Stereo pairs and 3-D reconstruction of (a) steel target, and (b) PVC pipes. Distance from scene shown by color coding. 27
16 Application of two different stereo techniques for 3-D scene reconstruction. 28
17 (a,d) Two coral objects, (b,e) occluding boundaries in sonar image, and (c,f) 3-D reconstruction of these boundaries, relative to the flat bottom surface, represented by red lines. Object size can be deduced from the contour reconstruction. 29
18 Templates of steel (left), aluminum (middle), and PVC (right) cylindrical targets at 6 orientations (for 0, 30, 60, 90, 120, and 150 [deg]) relative to the sonar. 30
19 Sample DIDSON images (a,b) before and (a',b') after application of CTR algorithm, from (a,a') pool and (b,b') marina data sets. 37
20 (a) Sample color image from a 1M-pixel optical camera. (b) Processed gray-scale image. While the processing removes a significant amount of clutter thus making it easier to identify the object under given conditions, optical images may not reveal much information under more turbid conditions. (c) Sonar image taken from DIDSON camera with $\alpha = 30$. 43
21 Acoustic images of the object O_Θ at angles $0 - 150$ in steps of 30 degrees. Note the variation in shadow geometry with α. 44
22 Preprocessed optical images of the object O_Θ at angles $0 - 150$ at steps of 30 degrees. In views (b,c), it is clear that object is cylindrical. However, view (a) could well be some rectangular object. This uncertainty is even more pronounced in view (d). 45
23 (a) optical image on acoustic image, and (b) vice-versa. ($\alpha = 30$ image set.) 46
24 Change in belief of target type θ_1 as more and more views are combined together. 48
25 (a) Stereo views in scattering media with magnified view of section over a cylindrical target before and after enhancement. (b) Overlaying a diagonal section covering all three objects to facilitate the comparison. 50
26 DIDSON images of various targets depicted in Fig. 3 – Cinder block, dumbbell, and steel, aluminum and PVC cylinders – at 3 orientations of 320, 340 and 360 [deg] relative to the sonar. 61

LIST OF TABLES

I Target detection results with steel object before and after CTR preprocessing. . . . 32
II Summary of detection results for steel target, before and after CTR preprocessing. . 33
III Classification results for a steel target before and after CTR preprocessing. 34
IV Summary of classification results for steel target, before and after CTR preprocessing. 35
V Estimating target orientation, before and after CTR preprocessing, for tank data. . . 36
VI Estimating target orientation, before and after CTR preprocessing, for marina data. 38
VII Summary of results in estimating target orientation, before and after CTR preprocessing. 38
VIII Estimating target position, before and after CTR preprocessing, for tank data. . . . 39
IX Estimating target position, before and after CTR preprocessing, for marina data. . . 40
X Summary of results in estimating target location, before and after CTR preprocessing. 40
XI Error analysis in locating the steel target (in units of [pix]), before and after CTR preprocessing. 41
XII Target Characteristics 42
XIII Mass Assignment Models of the Evidence Generated by Optical and Sonar Sensors. Note that, $A_1 = \{\theta_1, \theta_2, \theta_3, \theta_6\}$, $A_2 = \{\theta_4, \theta_5, \theta_6\}$, $A_3 = \{\theta_1, \theta_2, \theta_3\}$. 42
XIV Fused Evidence BoEs. Note that, $A_1 = \{\theta_1, \theta_2, \theta_3, \theta_6\}$, $A_2 = \{\theta_4, \theta_5, \theta_6\}$, $A_3 = \{\theta_1, \theta_2, \theta_3\}$. Highest and second highest masses in each BoE, representing those propositions with the highest 'support', are marked in **bold** and <u>underlined</u>, respectively. 48

List of Acronyms and Abbreviations

2-D	Two-Dimensional
3-D	Three-Dimensional
AUV	*A*utonomous *U*nderwater *V*ehicle
BBA	*B*asic *B*elief *A*ssignment
BoE	*B*ody *o*f *E*vidence
CCT	*C*onditional *C*ore *T*heorem
CTR	*C*ross*t*alk *R*eduction
CUE	*C*onditional *U*pdate *E*quation
DIDSON	*D*ual *ID*entification*SON*ar
DCR	*D*empsters *C*ombination *R*ule
DS	*D*empster-*S*hafer
EF	*E*vidence *F*ilter
FH	*F*agin-*H*alpern
ESTCP	*E*nvironmental *S*ecurity *T*echnology *C*ertification *P*rogram
FAU	*F*lorida *A*tlantic *U*niversity
FoD	*F*rame *o*f *D*iscernment
FS	*F*orward *S*can
LLS	*L*aser *L*ine *S*can
MAFR-TM	*M*ultiple-*A*spect *F*ixed-*R*ange *T*emplate *M*atching
MTA	*M*arine *T*owed *A*rray
PVC	*P*oly*V*inyl *C*hloride
ROV	*R*emotely *O*perated *V*ehicle
SERDP	*S*trategic *E*nvironmental *R*esearch and *D*evelopment *P*rogram
SNR	*Signal*-to-*N*oise *R*atio
UM	*U*niversity of *M*iami
UXO	Unexploded Ordnance

List of Symbols

Image registration

$\Re$	sonar range measurement
γ	sonar azimuth measurement
ϕ	elevation angle
λ	arbitrary scaling
$\mathbf{n}$	normal vector of ground plane (sea bed)
$\mathbf{p}_o$	point in optical image; $\mathbf{p}_o = (x, y, 1)$
$\mathbf{p}_s$	point in sonar image with coord. $\mathbf{p}_s = (x_s, y_s, 1)$; $(x_s, y_s) = (\Re \sin\gamma, \Re \cos\gamma)$
C	contour of object's cast shadow on ground plane (sea bed)
O	occluding contour of a 3-D object
$\mathbf{H}$	3×3 optical to sonar image transformation
$\mathbf{P}$	a 3-D point on scene surface; $\mathbf{P} = (\Re \sin\gamma \cos\phi, \Re \cos\gamma \cos\phi, \Re \sin\phi)$
$\mathbf{P}'$	point on Object Occluding contour
$\mathbf{P}''$	point on 3-D Object

MAFR-TM algorithm

$(\Delta m, \Delta n)$	template's pixel increment along rows and columns
$\Delta\theta_s$	orientation angle increment of source images
σ_S	standard deviation of source image
σ_T	standard deviation of template
∇_V	vertical gradient of source image
β	Kaiser window parameter
θ_S	orientation of source image
θ_T	orientation of template
$\Delta\theta_S$	orientation angle increment of source image (10 [deg])
$\Delta\theta_T$	orientation angle increment of template (10 [deg])
d	distance between target and shadow zones
d_{min}	minimum distance between target and shadow zones
(i, j)	indices of correlation function (between source image and template)
(i_H, j_H)	indices of points in target zone
(i_{max}, j_{max})	indices of maximum correlation between source image and any given template)

List of Symbols (continued)

(i_s, j_s)	indices of points in shadow zone
(i_T, j_T)	location of detected target over all templates
k_S	index of source image orientations
k_{Smax}	k_S for peak correaltion
k_T	index of template orientations
k_{Tmax}	k_T for peak correaltion
(m, n)	indices of template pixel positions
m_T	index of template material type
m_{Tmax}	m_T for peak correaltion
(p, q)	indices of Kaiser window function
$s(x, y)$	source image
$\bar{s}(x, y)$	mean intensity within 3×3 region of (x, y)
$t(m, n)$	template
$\bar{t}(x, y)$	mean template within 3×3 region of (x, y)
(x, y)	indices of source image position
(x', y')	modified indies of template position
A_H	target Zone
A_S	shadow zone
C	correlation function (between source image and template)
C_{ST}	correlation function for a particular template, source image and material type
$C_{ST,H}$	collection of points making up peak area of correlation
$C_{ST,N}$	mean correlation excluding peak area (noise equivalent in calculation of SNR))
C_{STNorm}	normalized correlation function (between source image and template
C_{STmax}	maximum correlation value
(I, J)	correlation window size
I_o	zero-order modified Bessel function of first kind
K_S	number of source image orientations (36 here)
K_T	number of template orientations (36 here)
(M, N)	length and width of a template ([pix]))
M_T	number of material types (3 here)
S	selected source image
T	selected template
T_d	minimum distance threshold
T_G	vertical gradient threshold
T_H	threshold of normalized correlation

(X, Y)	length and width of source image ([pix])
$(X_{ROI,T}, Y_{ROI,T})$	total horizontal-vertical pixels describing a template
$W(p, q)$	two-dimensional Kaiser window
	DS Theory and Formulation
$\Theta = \{\theta_1, \ldots, \theta_n\}$	FoD denoting total set of objects to be discerned
θ_i	object of interest; also referred to as a singleton
2^Θ	Powerset of Θ; support in DS theory is assigned to elements in powerset
$A \setminus B$	Singletons in $A \in 2^\Theta$ not in $B \in 2^\Theta$.
$\overline{A}$	$\Theta \setminus A$, i.e, complement of A.
$m_\Theta(A)$	DS theoretic BBA or mass assignment to $A \in 2^\Theta$.
$Bl_\Theta(A)$, $Pl_\Theta(A)$	Belief and plausibility of $A \in 2^\Theta$, respectively
$\mathfrak{F}_\Theta$	Core of a BBA, i.e., set of focal elements (propositions receiving non-zero mass)
$\mathcal{E} = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\bullet)\}$	BoE
$\mathcal{E}_1 \oplus \mathcal{E}_2$	BoE obtained by combining BoEs $\mathcal{E}_1$ and $\mathcal{E}_2$ via DCR
K	Conflict between evidence provided by $\mathcal{E}_1$ and $\mathcal{E}_2$ as captured within DCR
$\mathcal{E}_1[k] \oplus \mathcal{E}_2[k]$	BoE obtained by updating BoE $\mathcal{E}_1[k]$ with evidence in BoE $\mathcal{E}_2[k]$ at instant k via CUE
$\{\alpha[k], \beta(A)[k]\}$	Non-negative parameters of CUE
$\{m_\Theta(B\|A), Bl_\Theta(B\|A), Pl_\Theta(B\|A)\}$	Fagin-Halpern (FH) conditionals of $B \in 2^\Theta$ w.r.t. $A \in 2^\Theta$ for mass, belief, and plausibility functions, respectively
$\{\mathcal{E}_O, \mathcal{E}_S\}$	DS theoretic evidence obtained from optical and sonar sensors, respectively

Belief Theoretic Multi-Sensory Data Fusion for Underwater UXO Identification

Executive Summary

Enormous difficulties arise in effective detection and identification of unexploded ordnance (UXO), generally lying within harsh clutter-filled underwater environments, mainly due to rampant imperfections in the gathered data (i.e. missing, incomplete, uncertain and/or ambiguous entries). No single sensing technology can be both accurate and cost-effective for estimating even the density and spatial distribution of UXO objects, let alone identifying them [24]. An effective mechanism is required to enhance the discrimination capability by making use of various heterogeneous data from different sensing modalities and resolutions.

This report details the findings from the activities in a one-year collaborative project between engineering research groups from the University of Miami (UM) and the Florida Atlantic University (FAU). The goals were set to carry out an initial exploration of technical problems and technologies for the detection and discrimination of underwater munitions based on: 1) optical and acoustic monocular and stereo imaging; 2) a novel framework for fusing heterogeneous data that is suited to the UXO identification problem; and 3) potential deployment of a commercial ROV equipped with a suite of optical and acoustic sensors for data acquisition. The expected outcome was to identify the most promising directions of future research within the explored topics, where investment may be made for the necessary technical and technological developments.

This report is organized in several sections, highlighting the suitability and limitations of individual sensing technologies. These include high-frequency two-dimensional sonar video imaging systems for short-range high-resolution target mapping, which is expected to be a critical component of any multi-modality detection and identification system, and the potential in the integration of optical and sonar technologies.

I. Introduction

Aside from numerous unknown worldwide deep sea sites, munitions are present in millions of acres of ponds, lakes, rivers, estuaries, and coastal ocean areas adjacent to active and former Department of Defense (DoD) military installations. Chemical warfare material and unintentional detonations from various munitions, including the bombs, projectiles, mortars, grenades, and rockets can have a direct impact on human health and life. These threats also impact the marine environment, and eventually also lead to human injury or death from the exposure, toxic concentrations of chemical material washing ashore, and contamination of the food chain.

A SERDP/ESTCP three-day workshop in July-August 2007 was targeted at identifying critical technological needs and a strategic plan to direct investments in the area of detection and remediation of underwater UXO within 5-10 years. The workshop report calls for basic and applied research on phenomenology, sensor development, signal processing and other supporting technologies to address gaps in understanding, as well as system integration and demonstration with existing technologies to address gaps in capabilities [30]. In particular, some recommendations include establishing the appropriate combinations of fully integrated platforms, sensors, and navigation/positioning technology for different applications. Potential platforms include autonomous underwater vehicles (AUVs), remotely operated vehicles (ROVs), and towed arrays, while promising sensors include magnetometers, electromagnetic induction (EMI), magnetic, optical, sonar/acoustic, chemical sensors, and laser line scanning sensors. A combination and the integration of these technologies are necessary in order to handle the variability in bottom conditions, water clarity and depth, size and type of munitions of interest, whether they are buried or proud, etc. It is recommended that the initial test bed for the characterization and assessment of suitable technologies should be within waters shallower than 120 [ft], consistent with the Navy criteria for the inclusion of underwater sites in its munitions response program. The site should be located in relatively calm conditions with low potential for mobility over time, and readily accessible for mobilization of sensor systems. The bottom environment should be simple, flat, sand or hard mud, and with few obstructions, allowing the assessment of various mapping and reacquisition technologies in benign conditions. It is also conjectured that integrated systems are needed not only to determine footprints of contamination in a wide area survey, but also to perform detailed surveys for individual detected items, during a reacquisition process.

II. Project Overview

A. Background

There has been a long history in the application of sonar data processing for subsea applications, because acoustic energy can penetrate through silt, mud and various sources of turbidity. Sonar systems undoubtedly will play a critical role for UXO detection, and fundamental studies in the modeling of acoustic responses in the underwater environments are critical in understanding the significant variations in signatures with type/size of munitions, whether these are fully intact, distorted or broken into munitions-related scrap, filled/empty, buried, partially buried or proud.

Also, variations in the signature with the acoustic beam wavelength can provide the opportunity to discriminate between objects of different size and reflectance characteristics. However, conducting an extensive theory-based study and validation through controlled experiments were far beyond the scope of this one-year seed project. Instead, we proposed an empirical study, mainly targeted at assessing the potential advantages of color acoustic data from new generation of two-dimensional high-resolution forward-scan (FS) sonar video systems for the detection of UXO-like objects.

Where visibility allows, practically no sensing modality can match the information content from optical image systems, due to high resolution and data rate, as well as the very rich visual cues for target localization, discrimination and identification. The detection and classification capabilities can be further enhanced by making use of a large-area visual map by mosaicing a large number of images each covering only a small field of view. Here, the visual characteristics of the site (e.g., a debris field) provide information at scales larger than a single target and (or) small local region in the immediate vicinity. Furthermore, by collecting stereo data, one can generate 3-D topographical maps for selected objects of interest that may be identified in some region of the mosaic. Finally, a suitable framework is necessary that would allow the fusing of heterogeneous raw and processed data to achieve UXO identification effectively.

Our study has explored the use of visual cues in sonar imagery in two ways. In the *Multiple-Aspect Fixed-Range Template Matching* (MAFR-TM) approach, we merely rely on the acoustic signature in the neighborhood of each target, both the direct reflectance and the shadow regions, over multiple views from 0 to 360 [deg]. The goal to assess if the object type and material properties can be established from the neighborhood information. In our second approach, a data fusion method based on the *Dempster-Shafer (DS) belief theoretic framework* is explored for the integration with optical data for target detection and classification.

B. Objectives

This one-year collaborative seed project, carried out by a team of researchers from the University of Miami (UM) and the Florida Atlantic University (FAU), was aimed at exploring a number of technologies for the detection and discrimination of underwater munitions: 1) optical and acoustic monocular and stereo imaging; 2) a novel framework for fusing heterogeneous data that is suited to the UXO identification problem; and 3) potential deployment of a commercial ROV equipped with a suite of optical and acoustic sensors for data acquisition.

The project was designed to address some of the critical research and demonstration needs identified in the SERDP/ESTCP workshop report. Specifically, the research components included the assessment of a new generation of high-frequency 2-D sonar imaging systems for UXO detection, and a demonstration of their capabilities, as well as the exploration of a novel method for UXO discrimination and classification by the fusion of optical and acoustic data.

The objectives were 1) to explore the performance of each sensing modality alone; 2) to assess the impact of each sensor for the detection and discrimination within the integrated heterogeneous data fusion framework, and 3) to study how the strength(s) of each sensing modality may overcome the weakness(es) of the other modalities. Based on these, we can subsequently explore

in detail the establishment of an effectiveness mapping strategy and the development of UXO detection and classification systems that are robust against imperfections of the data collected in variety of underwater environments.

C. Proposed Plan and Testing

A comprehensive plan was formalized in the form of 1) theoretical developments within the multi-modal data fusion paradigm, and other specific methods to facilitate its application to optical and sonar data, e.g., the registration of optical and sonar data; 2) collection of data to support the testing of various theoretical developments.

In addition to the pool facility, where experiments with ground truth could be readily carried out, the plan was to select certain test sites in relatively calm conditions in shallow water depths, readily accessible for sensor deployment (5-40 [ft]). These would include both benign conditions - simple flat sandy or hard mud bottoms with few obstructions - and clutter that could potentially interfere with UXO detection. Among candidate locations, we proposed:

- The dock and marina facility at the FAU SeaTech Campus – easy to access, set up and conduct experiments from the shore, though limited to the water depth of about 7 [ft] at high tide.
- Key Biscayne National Park, Miami, Florida – marked sites accessible for repeat experiments at depths of 40 [ft], where we have previously conducted ROV experiments.

For targets, we selected objects of various sizes, shapes and material properties. These include various cylindrical pipes mimicking inert munition as well as clutter, e.g., a steel pipe with 4-5 to 1 length to diameter (L-to-D) ratio, and PVC pipes of 8-9 to 1 L-to-D ratio.

Our goal was to place these targets within 2-3 areas of roughly 300-400 [sqm] in size, to be imaged over time. Each, simulating a region of interest (ROI), was assumed to have been identified in an earlier wide area survey (WAS). By carrying the data collection several times within the first two quarters of the project, we hoped to account for variability of the environmental conditions in our experiments.

Finally, the sensor deployment within the ocean site(s) was planned onboard a commercial custom-built tethered remotely operated vehicle (ROV), recently purchased by the UM team. The standard equipment includes: 1) pair of 1 M-Pixel digital optical stereo cameras; 2) analog video-resolution navigation camera, 3) a DIDSON sonar video camera operating at 1.1/1.8MHz[1]; 4) Imagenex[2], a 3-D scanning sonar operating at 675kHz with a maximum range of 50m; 5) Trackpoint-III[3] acoustic tracking/positioning system to establish the ROV position. Additionally, the ROV electronics was designed to allow for the installation of a blazed-array BlueView[4] DF-900/2250 sonar, operating at 0.9/2.25MHz, in conjunction with the DIDSON data.

[1]Trademark of Sound Metrics Corp., Lake Forest Park, WA.
[2]Trademark of Imagenex Technology Corp., Port Coquitlam, BC, Canada.
[3]ORE Offshore, West Wareham, MA.
[4]BlueView Technologies Inc., Seattle, WA.

Fig. 1. FAU's outdoor tank facility in Boca Raton (left) and marina at Sea Tech Campus (right).

D. Revised Plan

During the course of this project, certain difficulties and obstacles were encountered, mainly dealing with data collection. Overcoming some within the time frame could not be achieved without compromising some of the major objectives, and therefore certain adjustments to the original scope became necessary.

First, software had to be developed for the effective use of our imaging systems on the ROV; namely, the *"concurrent acquisition of data (on the same run)"* although not necessarily in synchronous fashion. Also, the recorded data had to be displayed live to ensure proper operation of the entire system. Unfortunately, the personnel recruited to work on the technical aspects of the proposed project, namely image processing and data fusion tasks, were not trained for the specialized tasks of developing a real-time acquisition and display system. In particular, we determined that the objectives of the proposed work could be readily met by stand-alone deployment of our systems, without the ROV. The final plan, discussed with and approved by the SERDP Technical Program Manager, became to limit the scope of the experiments to the FAU outdoor pool and marina facilities; see Fig. 1. We simply built individual cables for various imaging systems.

Next, we were faced with high turbidity and thus very poor visibility due to adverse environmental conditions during a relatively long period of several months at the FAU marina. A visibility of about 8-10 feet was necessary in order to collect optical data for site photomosaicing, 3-D target reconstruction from stereo images, as well as optical and sonar image registration to test our data fusion methods. As a result, we proposed and were granted a plan revision in order to carry out these tasks solely with the data collected in the FAU outdoor pool.

The small UM indoor pool was also used for the calibration of the sonar equipment, and the collection of some optical and acoustic data to test certain registration and sonar processing algorithms.

III. Overview of Technical Work and Accomplishments

This study was focused on exploring the potential application of several technologies for the detection and localization of intact UXO objects on the sea floor. The major accomplishments during the project are as follows:

- Data Collection:
  - Design and construction of a sensor rig to deploy multiple optical and acoustic cameras for pool and marina data collection;
  - Collection of a large database of sonar video clips with DIDSON and BlueView systems in the outdoor pool and marina at the FAU facilities;
  - Collection of some optical data in the outdoor pool for mosaicing, optical and sonar image registration, and data fusion testing.
- Application of existing optical technologies and performance assessment
  - Photo-mosaicing for scene interpretation (e.g., a debris field);
  - stereovision for 3-D target shape reconstruction
- Theoretical and algorithmic developments:
  - Optical and acoustic registration;
  - 3-D object contour reconstruction by acoustic shadow analysis;
  - MAFR-TM sonar processing technique for target detection and classification;
  - DS-based data fusion
- Ground truth experiments and technology assessment;
- Performance of MAFR-TM algorithm;
- Integration of optical and sonar data with DS-based data fusion algorithm

One conclusive outcome of this investigation is that high-frequency 2-D FS sonar video systems have high potential for UXO detection and classification. Having come to commercial market mainly as a visualization tool for environments with limited and/or poor visibility, these systems are undergoing continuous improvement that should enhance target discrimination capability based on material properties. Imagery with higher resolution and target details is expected with the emergence of more advanced hardware technologies and signal processing techniques, as well as the implementation of design improvements.

Acquiring data under identical imaging conditions has allowed us to preform a direct comparison between two existing technologies, DIDSON and BlueView systems. Not surprising, DIDSON provides better target details due to the higher spatial resolution associated with narrower horizontal beam widths. Because of our time limitations, we focused our investigation primarily on the application of our methodologies to the data acquired with a DIDSON system.

We have developed the Multiple-Aspect Fixed-Range Template Matching (MAFR-TM) technique mainly to assess the information content of the FS sonar image in the neighborhood of each target, and how this would vary with the viewing direction. In this method, we make use of both the direct reflectance and shadow regions in determining the object type and material properties. Additionally, preprocessing the data for cross-talk reduction has suggested higher discrimination

(a) (b)

Fig. 2. Camera assemblies for (a) pool and (b) marina deployment.

capability between metallic and non-metallic targets, even though our experiments were based on matching these preprocessed images with target templates constructed from unprocessed data.

Our novel method for constructing a target's 3-D occluding contour from overlapping optical and sonar images was aimed to address the adverse impact of visibility limitations on contrast and target details in optical images. First, the optical and sonar image positions of a minimum of three features on the sea floor – potentially small markers with distinct acoustic and optical reflectance properties – allow us to determine the orientation of the bottom surface relative to the imaging platform. The balance of the computations rely solely on the visual information in the sonar image, namely, the object's cast shadow region and boundary. The immediate application is to label occluding contours as straight or curved, in discriminating between manmade UXO and clutter/natural objects, respectively. Although the shadow shape can change due to imperfect boundary detection, clutter and various sources of occlusion, including sediment coverage, the ambiguity may be partially resolved by imaging the target from various viewing angles. More importantly, the DS theoretic framework is just the right paradigm to handle such ambiguities that arise from imperfect data.

A distinct technical contribution of our work is a data fusion method based on the *Dempster-Shafer (DS) belief theoretic framework* for the integration of data from multi-modal imaging systems, as well as information from heterogeneous sources and other sensors. In our work, the scope was restricted to evidence obtained from optical and sonar image data and the fusion of this data via the widely used DCR and a recent DS theoretic evidence updating technique. In light of the study we have undertaken and the lessons learned from it, we have identified several promising avenues of research and exploration that are described in the last section.

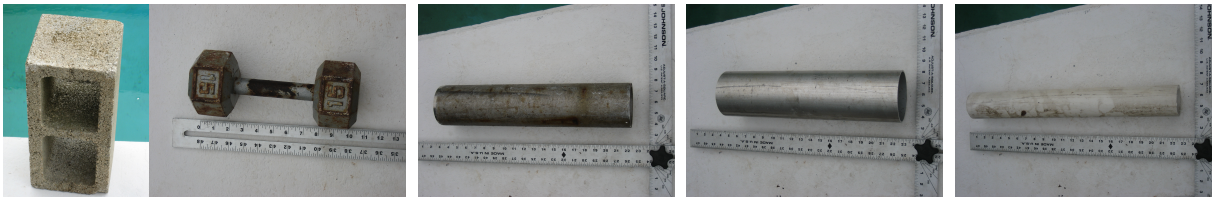


Fig. 3. Various targets with different shapes and reflectance properties; cinder block, dumbbell, as well as steel, Aluminum and PVC pipes

IV. Project Tasks

A. Data Acquisition

A significant part of the project was dedicated to data acquisition activities, resulting in a rather large volume of recorded data, organized by time, sensor and location. Part of our effort was expended to improvise and implement the new plan in place of the original proposed strategy. As stated, ocean data were to be obtained by deploying the UM commercial ROV in depths of 40 [ft] or less within selected sites in the Miami Beach and Key Biscayne areas. While the platform is designed to carry sonar and optical cameras, software for *concurrent data acquisition with various systems* had to be developed. This could not be accomplished based on the expertise of the personnel recruited for the image processing tasks, and thus suitable deployment strategies for outdoor pool and marina sites had to be adopted.

For collecting the data in the 5 [m] deep FAU outdoor pool, two camera mounting assemblies were designed and built; see Fig. 2. These were to support optical cameras, a single DIDSON video camera and up to two BlueView sonar cameras, for both the pool and marina deployment.

The various sonar data were acquired with two different video camera types, namely, the BlueView Technologies DF900/2250 and the Sound Metrics Didson US300. The salient features of these sensors are:

- BlueView DF900/2250 operating at 900kHz and 2.25MHz: maximum range of 60m; 256 beams, each with horizontal and vertical widths of 1 [deg] and 20 [deg], respectively, spaced at 0.18 [deg]; range resolution of 2.54 [cm]; and an update rate of up to 10Hz.
- Sound Metrics DIDSON US300, a dual-frequency sector-scan sonar operating at 1.1 and 1.8MHz: maximum range of 40 [m] (at 1.1MHz); either 48 (at 1.1MHz) or 96 (at 1.8MHz) beams of 0.3 [deg] in horizontal and 14 [deg] in vertical widths. These are acoustically focused with a field-of-view of roughly 28.8 [deg], and update rate of up to 21Hz.

Fig. 3 depicts a number of metallic and non-metallic targets; namely, a 15 [lb] dumbbell, a cinder block, and cylindrical tubes made of steel, aluminum, and PVC. The acoustic signatures of these objects were obtained in the pool from multiple viewing directions at 3-4 frequencies in the 0.9-3.0MHz range. The main purpose of imaging was to test the functionality of the sonar camera under controlled conditions. Specifically, these objects were used to assess the sonar image quality for objects with different shape and reflectance properties, and to determine if these objects were visually identifiable from their images.

Sonar views of each of these targets at three orientations, acquired by a DIDSON camera are given in Fig. 26 in Appendix B. Based on this initial assessment, the UXO-like objects (steel,

steel pipe Diameter & Length 4 [in]×18 [in]	aluminum pipe Diameter & Length 2.5 [in]×22.25 [in] L	PVC pipe Diameter & Length 2.75 [in] × 21.5 [in]

(a) (b) (c)

Fig. 4. Three cylindrical targets used in the construction of the source image data base for the testing of the MAFR-TM technique

aluminum and PVC cylinders) were used in the detection and classification tests in the pool and in the marina. Furthermore, although numerous sonar data sets have been collected in both the pool and the marina, the processing for target detection and classification has been limited to the data with the highest resolution; namely, the images from a DIDSON with 0.3 [deg] horizontal beam width, operating at 1.9 [MHz].

To explain, we have noted from our data that the BlueView images with lower azimuth resolution do not provide sufficient target details for the application of our methods. However, it is also fair to say, based on the BlueView data we have collected, that this sonar can be an effective tool to discriminate between metal and non-metal targets. To be precise, blob-like appearance of small objects is insufficient to obtain detailed structural information about a target shape, from either intensity profile or cast shadow geometry that our two methods exploit. However, metal objects have a significantly stronger return compared to non-metal targets for most viewing angles. As we present some of our data, this can be assessed qualitatively from side-by-side DIDSON and BlueView images, acquired under similar viewing geometry.

We had allocated two Mpix analog progressive JAI CV-M77 cameras, with 100 [ft] cable, exclusively for pool and marina deployment during the entire project. Unfortunately, these cameras malfunctioned during one of the initial marina data acquisition operations. In their place, we had to make use of two 1 [Mpix] pointgray digital color cameras, that are part of the UM ROV imaging cameras. Thus the 600 [ft] ROV tether had to be hauled with the cameras between the FAU outdoor pool (Boca Raton campus), the FAU marina at the Sea Tech Campus (Dania Beach, FL), and the UM indoor pool (Coral Gables, FL).

1) Tank Data: We have recorded a large amount of data in both the UM indoor and FAU outdoor tanks; Appendix B depicts samples from this extensive data base. Some of the data sets have been utilized to test and demonstrate the various aspects of our technical work. Due to the limited size of the UM indoor pool, the use of recorded data was primarily to calibrate the optical and sonar cameras, and to carry out the initial tests for specific algorithms; e.g., optical and sonar image registration. In addition to these, various sonar and optical images of certain targets were recorded in the FAU outdoor pool.

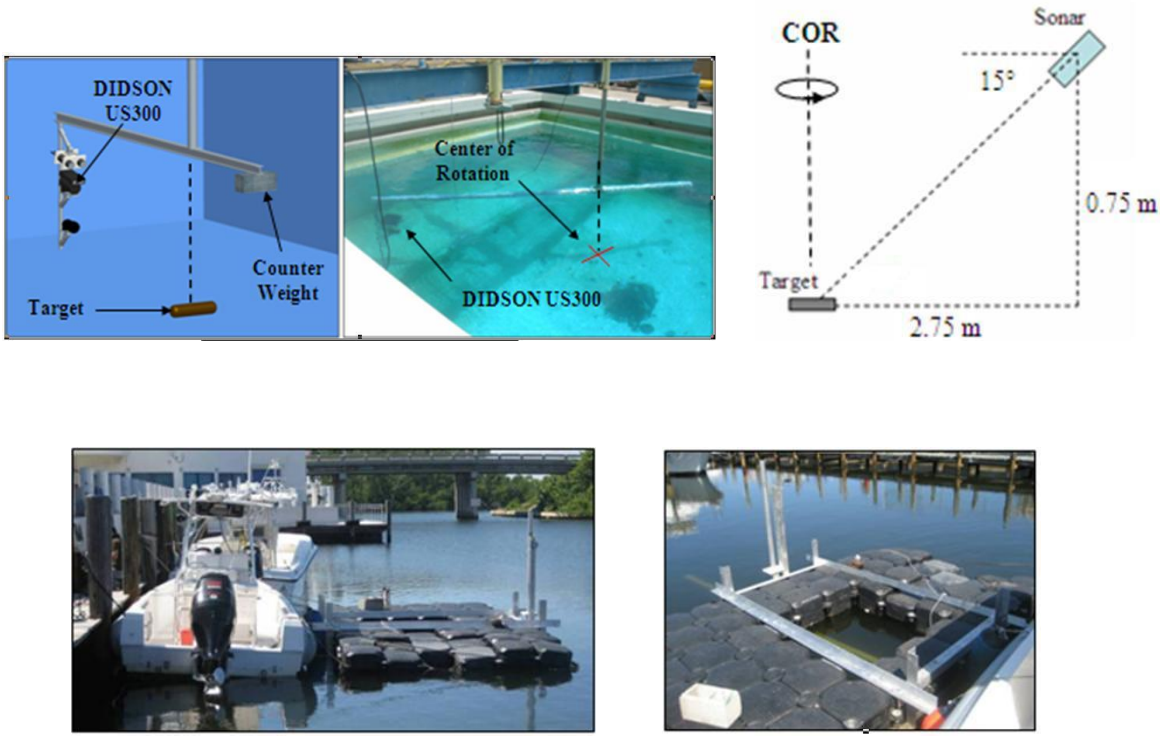


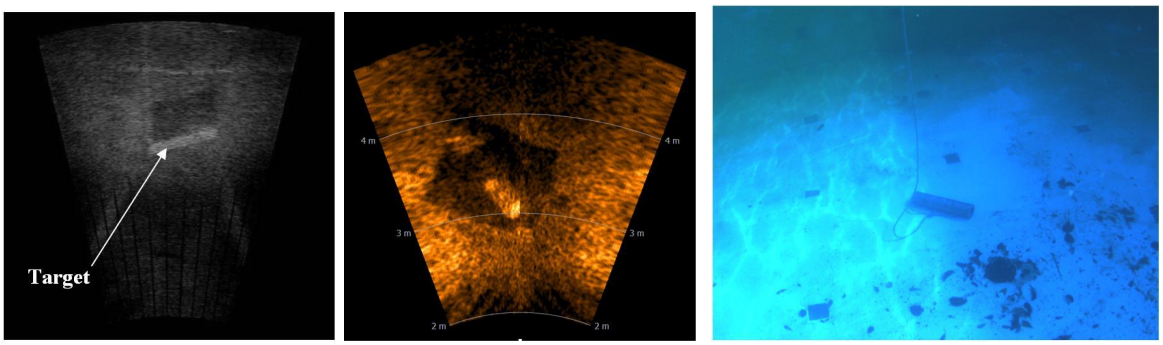
Fig. 5. Tank (top) and marina (bottom) setups.

Optical images were recorded in the FAU outdoor pool to build a site mosaic, to generate 3D topographical maps from stereo data, and to use with the sonar data in the application of our DS-based data fusion algorithm.

Additionally, optical and sonar images of targets with different material properties were recorded at different orientations relative to the sonar. The sonar component, namely the DIDSON frames, form the data base of source images for the application of MAFR-TM technique, which is described in section IV-B4. Our sonar data sets were acquired under similar settings over a period of approximately 9 months. All together, they comprise an extensive record of sonar imagery for cylindrical targets with different material properties and broad spatial and temporal resolutions.

Fig. 4 depicts three cylindrical targets, made up of heavy gage steel, aluminum and polyvinyl chloride (PVC) with a minimum length to diameter ratio of 4:1. The sonar and optical images in this data set were recorded at a fixed target distance of 3 [m] at 36 different orientations, from 0 [deg] - 350 [deg] in 10 [deg] increments with respect to the camera. This comprise a total of 108 sector-scan sonar as well as optical images per target, with no obstructions, occlusions, and very limited bottom roughness.

To collect these images, the imaging system was mounted on the assembly shown in Fig. 2. Secured at one end of a 5 [m] cantilever beam, approximately 3 [m] from the center of rotation,



(a) (b) (c)

Fig. 6. Sample data of steel cylinder acquired in outdoor pool with (a) DIDSON and (b) BlueView, and (c) with an optical camera.

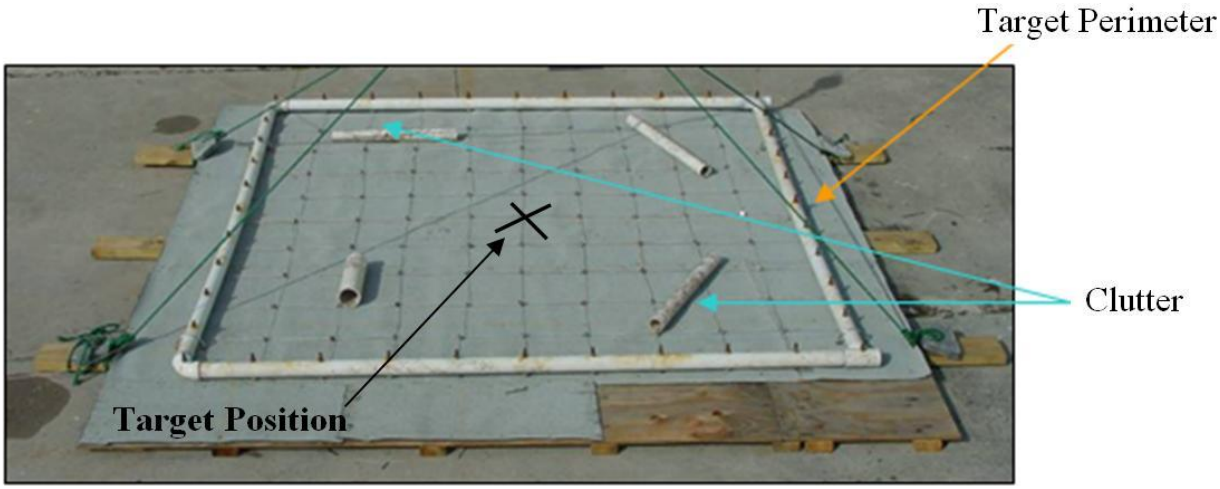


Fig. 7. Setup for target imaging with clutter in both outdoor tank and marina.

the target is imaged at a 15 [deg] viewing angle; see Fig. 5 (top). Fig.6 depicts a sample set of DIDSON, BlueView and optical images.

For testing the MAFR-TM technique, other images were recorded after sparse false targets (clutter) were added to the scene, for a more realistic aspect to the data; see Fig. 7. To maintain the same geometry for pool and marina testing, a 2.5 [m] $\times$ 2.5 [m] compact portable fixture was designed. It consists of a square PVC border attached to a plywood base. By running mono-filament lines across in both directions, we constructed a grid to which the target and clutter can be attached. Figs. 8(a-d) are sample BlueView and DIDSON images recorded in the outdoor tank.

2) SERDP/ESTCP Data: We had originally proposed to obtain two data sets from ESTCP under projects SERDP UX-1322 and ESTCP-200324 [1], [19], for the initial testing of our data fusion algorithm. These are 1) magnetometer from DUCK data set collected during MTA survey in parts of the Currituck Sound, adjacent to the former Duck Naval Bombing Range North of

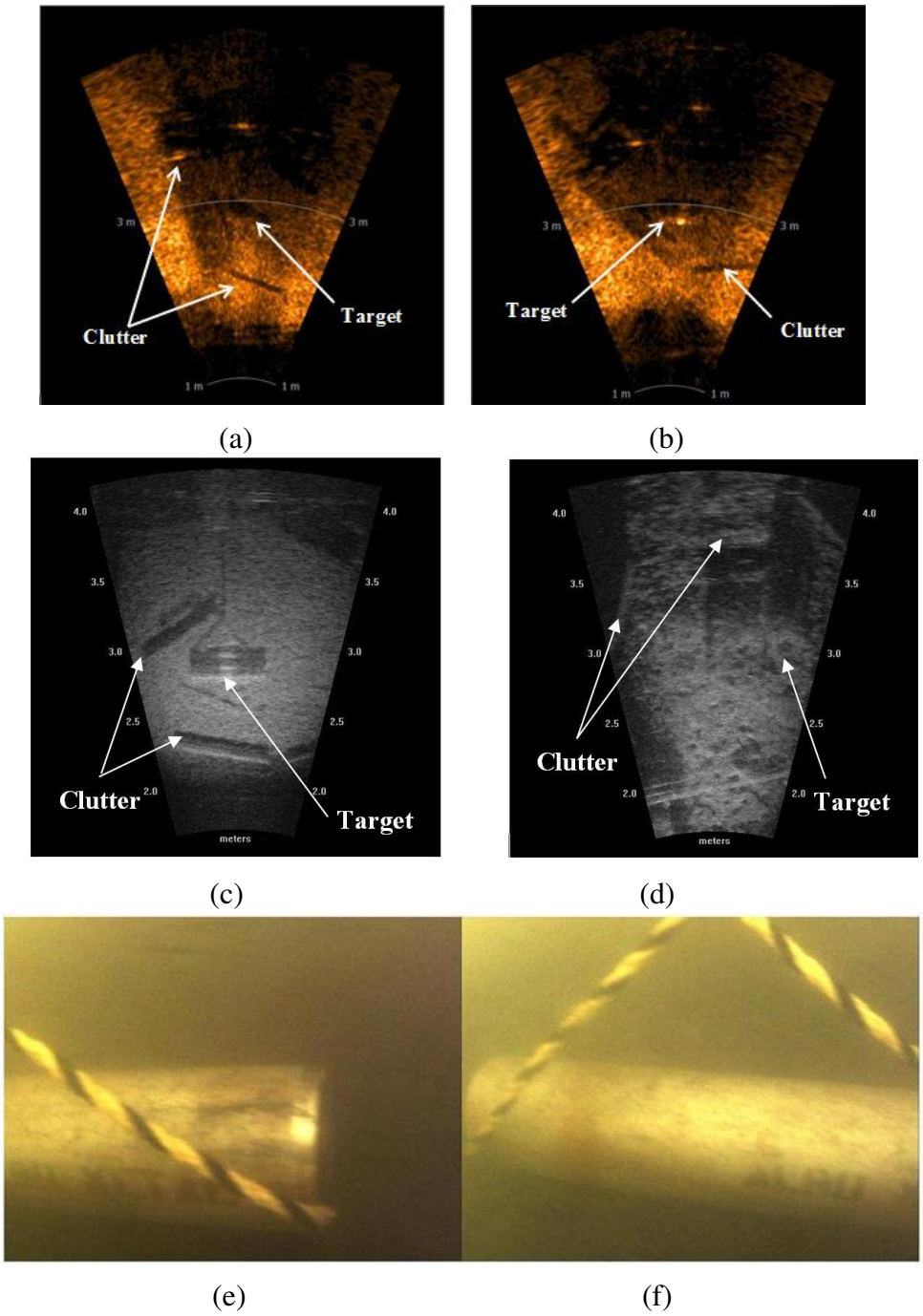


Fig. 8. Sample images acquired in FAU marina, with (a,b) BlueView at 120 and 160 [deg] and (c,d) DIDSON at 0 and 120 [deg] relative to the target. (e,f) optical stereo pair at about 2 [ft] distance.

Duck, NC [19]; 2) Pueblo data set [1]. We acquired both sets, and after some initial efforts, we were able to decode the first, but faced with problems in decoding and file formats of the latter.

Additionally, having no ground truth, we assessed this task as NO-GO. Instead, we used some data from the UM tank for our initial tests.

3) Marina Data: A large part of our data set was obtained at the dock facility of the FAU SeaTech campus. The portable fixture supporting the target and clutter objects was moved to the shallow waters of the FAU marina, a maximum depth of 2.2 [m] at high tide. A soft muddy bottom, turbid waters, and some target aging with biological growth provided a realistic testing environment, where detecting the target would often become difficult even by human eye. In particular, we encountered more serious variation from the sediments than from aging. Specifically, mud slowly would deposit on the portable target frames in the marina. The partial coverage of the target would affect its detection based on shape characteristics at high frequencies. Also, the temporal variations in environmental conditions – currents, waves, etc. – was responsible for changes in the sediment amount and distribution, thus resulting in variations in the target acoustic signature recorded at different times.

In order to acquire the data in the marina, the cantilever beam was secured to a floating barge that was fixed in place next to the FAU research vessel R/V Lee; see bottom row in Fig.5. Also, a more compact sensor mount was built to deploy the sonar and optical cameras, as shown in Fig.2(b). As in the pool data, the target alone serves as the template (reference). The source images of the target with clutter (for testing) were acquired at a fixed range of 3 [m] at 36 different orientations of 0 [deg]- 350 [deg] in 10 [deg] increments. Fig. 8(a,b) depict two sample images acquired with a BlueView sonar at 120 and 160 [deg], and with the DIDSON system at 0 and 120 [deg] angles relative to the target, respectively.

While we were able to record a large set of acoustic data in the marina during a 9 month period, we have little success with optical video/stereo imagery, facing turbid conditions due to adverse weather through most of late Fall'09 and Spring'10 seasons. Fig. 8(e,f) depict a sample stereo data at only a couple feet target distance. In comparison, we had planned for a reasonable visibility at about 6 [ft] to record useful data for our photo-mosaicing and stereo reconstruction algorithms. Due to these adverse conditions, we moved our optical data acquisition set up to the outdoor pool at the FAU Boca Raton campus, at roughly the half way point of our project.

B. Data Processing

As described, we have collected some limited data in the UM indoor water tank for the initial testing of our algorithms, an extensive set of pool and marina acoustic data, as well as optical data in the FAU outdoor pool.

1) Test Site Mosaic Generation: We were involved in the development of the initial automated system for the generation of high-resolution mosaics from thousands of images, originally targeted for the monitoring and health assessment of coral reef communities over 100's [sqm] in area [27]. Such a large area map can play a significant role in UXO localization and identification, since decision making can be made based on large-scale visual characteristics of the site; e.g., a

debris field, rather than just a single target and (or) a very small local region in the immediate vicinity.

As we demonstrate, the quality and accuracy of a mosaic is directly tied to the visual information content of the scene surfaces. Ideal characteristics include strong surface markings and texture, reasonably uniform illumination over the region being imaged in each frame, and small topographical variations relative to average distance to the scene.

2) 3-D Reconstruction: Binocular stereo imaging has been established as a popular approach in reconstructing 3-D terrestrial maps of terrains and objects, from pairs of overlapping 2-D optical images. It has also been employed in underwater environments with good visibility, however, it becomes constrained or practically ineffective with increasing turbidity. In particular, while the matching of certain distinct features in two views gives their 3-D positions, this falls far short of constructing a dense 3-D map as a target structural descriptor.

In addition to the generation of a mosaic, our goal was to explore if 3-D structural information can potentially enhance object characterization and classification, and (or) resolve potential ambiguities. Rather than building a 3-D topographical map of the entire site from stereo images, which is possible but computationally demanding, one can effectively construct a high-resolution 3-D map over regions of the mosaic where potential targets may exist with higher probability.

3) Optical and Acoustic Image Registration and 3-D Reconstruction: A potential UXO target is not always detectable in every view and/or imaging modality, however, locating it in some view(s) and one of the two image modalities can significantly enhance the detection in other views and (or) modality. In our application, we are interested in the registration of 1) images within the same modality, namely optical or sonar images, e.g., to build a composite mosaic, acquired as the sensor platform navigates over the a potential debris field; 2) multi-modal data, i.e., optical with sonar, in order to fuse *"opti-acoustic visual cues"* for target detection and classification.

Image Registration: Optical and sonar cameras each have different field of view and operating range. Thus, it is likely that the sonar and optical images to be registered are not recorded simultaneously at the same sensor platform position, but rather at different times and platform positions, as we navigate over the site. This suggested that knowing the platform position at image acquisition times can be useful to address the registration problem. In particular, for optimal data fusion through space and time, the estimated vehicle motion and/or position must be highly precise. In addition, the accurate estimation of the platform trajectory is important for the effective execution of a local search, revisiting the same site and other navigation related capabilities. However, fundamental problems in vehicle navigation are too complex to be addressed adequately within the scope of this one-year seed project.

Various methods can be applied to register data from the same sensing modality that may have been obtained at different times and (or) sensor positions [23], however, it is not always easy to do the same for images from different modalities due to dissimilar nature of the image

formation physics.

The multi-modal image registration problem can be simplified by placing and making use of distinct markers on the sea floor. By locating them and establishing the so-called *"opti-acoustic correspondences"* – i.e., the association between the optical and sonar image positions of corresponding pairs – we can attempt to estimate the transformation between a pair of overlapping sonar and optical images. In practice, employing markers with distinct optical and acoustic reflectance characteristics can be adopted as an effective strategy to simplify both the intermodal and intramodal registration problems in *difficult environments*, where the identification and matching of natural features in two views is a formidable task.

The underlying mathematical model of the transformation between overlapping sonar and optical images is relatively complex, and depends on both the relative pose of the two cameras and the scene geometry. For a planar scene (e.g., the sea floor), it simplifies somewhat and can be approximated by a fixed up-to-scale linear transformation:

$$\lambda \mathbf{p}_o = \mathbf{H}\mathbf{p}_s, \tag{1}$$

where $\mathbf{p}_o = (x_o, y_o, 1)$ and $\mathbf{p}_s = (x_s, y_s, 1)$ are the homogenous coordinates of corresponding optical and sonar image points, $\mathbf{H}$ is a 3×3 transformation matrix, and λ is a feature-dependent scale factor. The sonar image is defines by $(x_s, y_s) = (\Re \sin\gamma, \Re \cos\gamma)$. Once $\mathbf{H}$ is estimated from the data, it can be applied to any sonar image point $\mathbf{p}_s$ to determine the corresponding optical position up to the scale λ. The absolute position (x_o, y_o) is determined by dividing the result $\mathbf{H}\mathbf{p}_s$ by the third component λ. Conversely, the sonar image match of an optical point is similarly determining by applying the inverse transformation $\mathbf{H}^{-1}$. We can fix $\mathbf{H}$ with a minimum of 4 opti-acoustic correspondences of features on the sea floor, giving us 8 linear equations in 8 up-to-scale unknowns of the transformation. The registration is accurate over the sea bed, and sufficient for smaller targets; it is somewhat approximate for larger objects.

As a selected example, consider Figs. 9(a-c) depicting the sonar and optical data collected in the UM indoor water tank, and the superposition of the same two images after registration in the red and blue channels of a color image. Suppose the goal is to locate the metallic cylinder and two coral rocks between the two long PVC pipes. In this example, the images can be registered on the plane defined by points on the top surface of the two PVC pipes, about 2 inches above the tank floor. The rough registration of these other 3 objects can often be sufficient to establish the association between their sonar and optical views, track them over multiple views, and fuse the multi-modal visual information for target classification.

As we note in (c), the registration of optical and acoustic images based on the above linear model leads to a good alignment over the flat floor. Although somewhat rough, the alignment of 3-D objects is adequate to restrict the search for each object, missed in one modality, to a small region around the position where it can be detected more readily in other modalities. We elaborate on the process next.

<u>*Object Segmentation:*</u> Many segmentation and clustering algorithms can be applied to locate potential targets of interest in an image, depending on a priori knowledge about the number

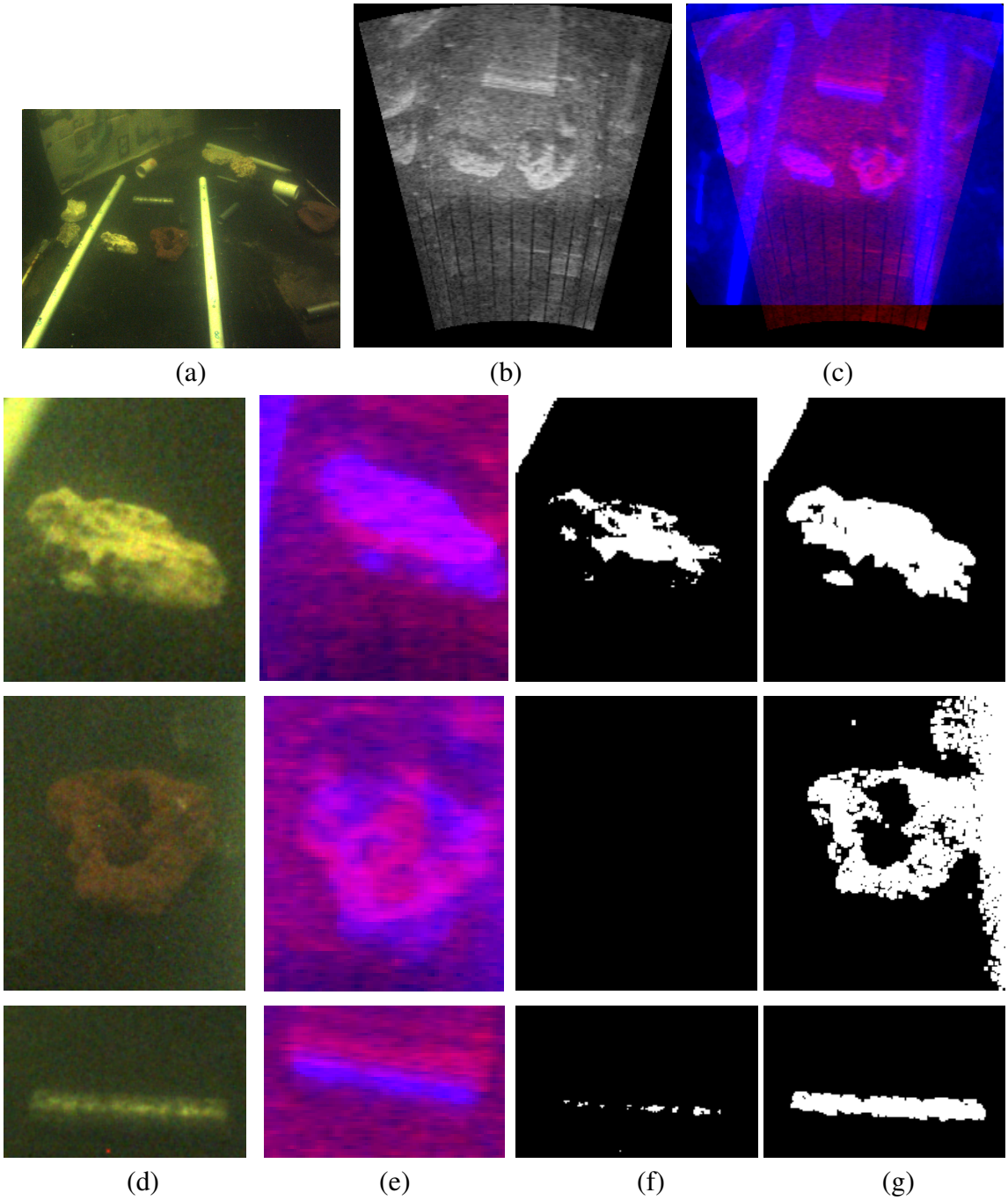


Fig. 9. (a,b) Overlapping optical and sonar views; (c) registered over the plane defined by the two PVC pipes. (column d) selected 3-D targets with different visual characteristics in the optical image and (column e) registered views. In absence of a priori information about number and visual properties of these targets, segmentation results is rather poor by applying an optimum global threshold over the entire image based on simplistic bimodal distribution model (column f)). Despite coarse image alignment of these 3-D targets, locating each in either/sonar image modality enhances ability to locate it in other/optical view, by adjusting detection parameter(s) optimally based on intensity distribution within region around each target (column g).

and each object's visual characteristics. The segmentation based on the simplistic bimodal distribution model applies well with one or more targets of similar mean intensity within a distinct background. It requires the selection of an optimum threshold for the entire image. The results in column (f) shows the deficiency of the model, where the objects of interest each can have a different mean intensity. The metal pipe and the brighter coral are partially segmented, while the darker coral is lost in the background. However, the model can produce good results when applied to a a local region around each object in one imaging modality, based on the knowledge of its position in the other modality, and only a rough alignment between the two views over the target of interest, as demonstrated in column (g).

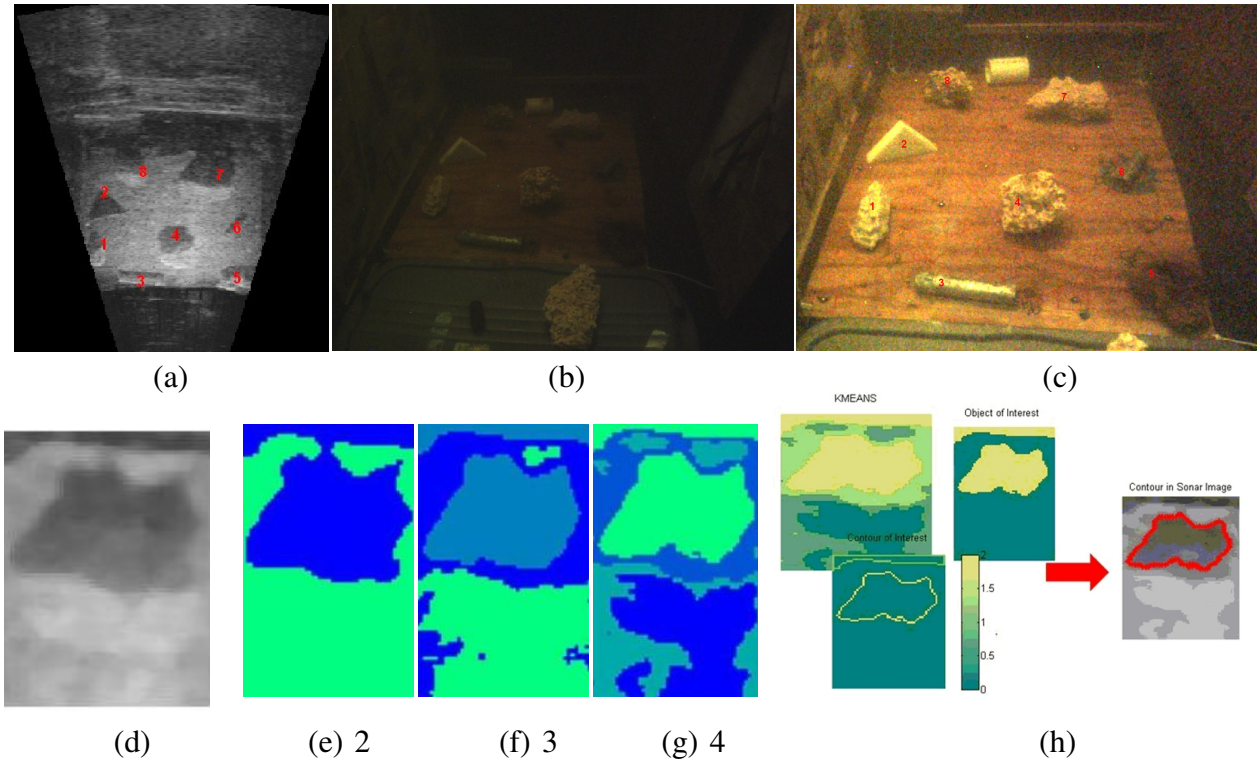
The k-means clustering, an effective unsupervised learning algorithm, is a generalization for segmenting an image into k clusters, each corresponding to the background and one of several targets with a unique mean intensity. This can also apply to UXO targets, where the visual appearance can vary significantly in both the optical and sonar images due to environmental conditions, occlusion and sediment coverage, variations in reflectance with viewing geometry, shading and cast shadow, etc. The goal is to segment the image into regions corresponding to all instances of varying UXO appearances, other natural objects, as well as the sea floor. The main complexity in applying the k-means method is that the performance can be sensitive to the number of clusters, and the distribution of each object intensity.

Figs. 10(a,b) depict a another sample pair from our water tank data sets, with overlapping sonar and optical images of targets with different shapes. In addition to the original optical image (b), we have enhanced the same view in (c) after enhancement for better visualization. In the sonar image, we have numbered various objects. Applying the k-means algorithm to object 7 with 2-4 clusters, we obtain different results. As noted in (e), a bimodal distribution model assuming 2 clusters can produce very good results when applied to a small region around the potential target. While suitable when applied to a larger or the entire image region, assuming a larger number of clusters can lead to over segmentation, over the confined search window; see (f) and (g).

Reconstructions of 3-D Occluding Contour: Binocular stereo imaging is a popular approach for reconstructing 3-D terrestrial terrains and objects, from pairs of overlapping 2-D optical images. It has also been employed in underwater environments with good visibility, however, it becomes constrained or practically ineffective with increasing turbidity. In particular, while the matching of certain distinct features in two views gives their 3-D positions, this falls far short of constructing a dense 3-D map as a target structural descriptor.

Next, consider an overlapping pair of optical and sonar images, which can be treated as multi-modal stereo views, as in Figs. 9(a,b) and 10(a,b). In [21], we have proposed opti-acoustic stereo imaging as a novel paradigm for 3-D object reconstruction. However, this method also requires sufficient visibility to image targets of interest with reasonable details in the optical image, but more importantly, we have to address the very complex opti-acoustic correspondence problem, namely, establishing the match of each sonar feature in the optical view, and vice versa.

In this work, we have developed a new approach that exploits the information is sonar images,



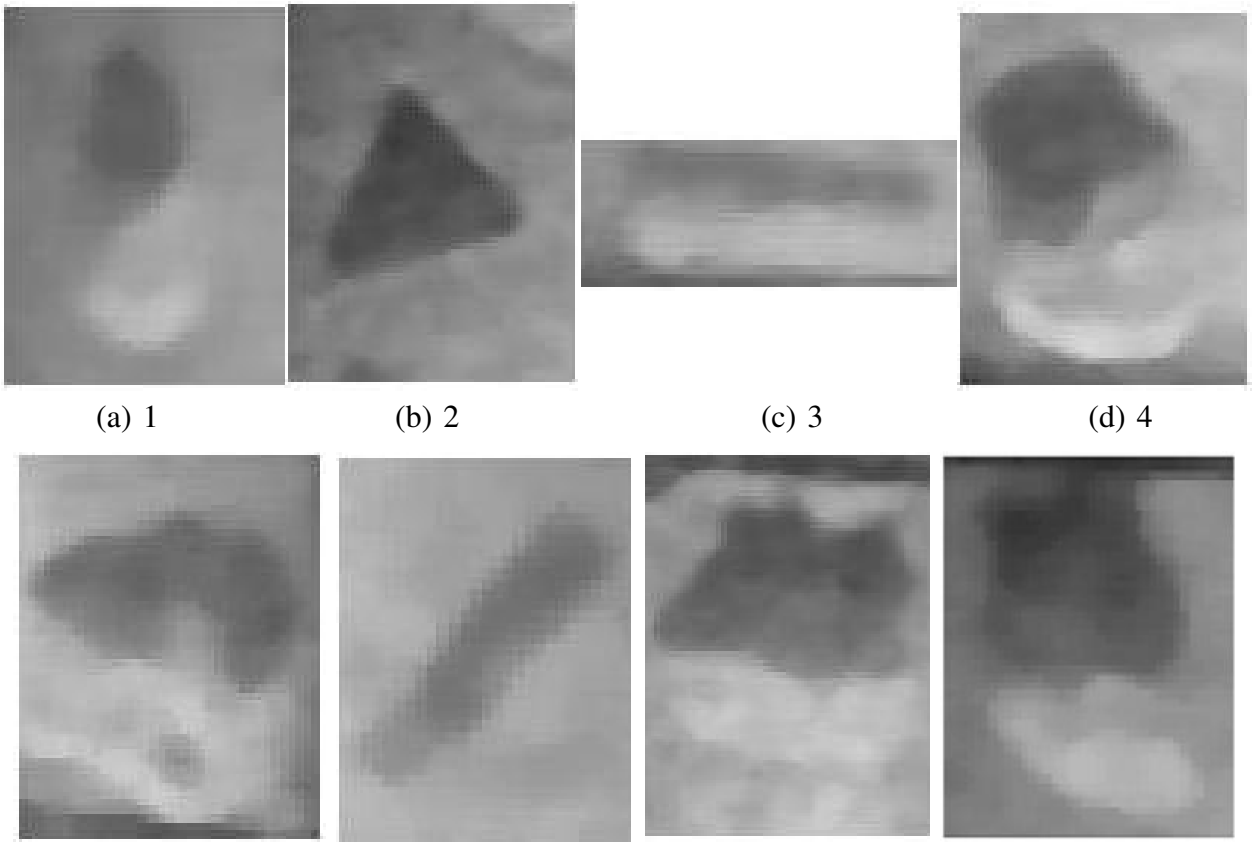
(a) (b) (c)

(d) (e) 2 (f) 3 (g) 4 (h)

Fig. 10. (a-b) One sample sonar and optical images from a data set acquired in UM indoor tank with clear water. (c) enhanced optical image for better visualization. (d-g) A particular object (labeled 7), and segmentation results by apply k-means with 2-4 clusters. Bimodal distribution for 2 clusters can produce very good results when applied to a small region around the potential target. Region can be over segmented when a larger number of clusters is assumed.

while minimally relying on optical views to reduce sensitivity to visibility limitations. We make use of the same feature correspondences employed in the registration of optical and sonar views, as well as the shadow region for each object in the sonar image; see Fig. 10(h). Utilizing these, we construct the object's 3-D occluding contour. Fig. 11 depicts different objects from the scene in Fig. 10, each taken from one of several sonar views in the set. As one can note, the object's cast shadow on the background sea floor surface can encode valuable information about an object's shape and its occluding contour in high-frequency 2-D sonar images.

Recall that we discussed employing a minimum of 4 seabed feature correspondences to register the sonar and optical views. We can apply the opti-acoustic stereo solution to a minimum of 3 such matches, denoted $\{\mathbf{p}_s, \mathbf{p}_o\}$, to determine their 3-D positions $\mathbf{P}$ [21]. Now, these allow us to determine the plane $\mathbf{n}$ of the sea floor. Knowing the plane, and given any sea floor feature in the sonar image, we can determine its 3-D position: $\{\mathbf{p}_s; \mathbf{n}\} \rightarrow \mathbf{P}$. More precisely, the sonar position gives the range $\Re$ and azimuth γ measurements, and we can calculate the elevation angle ϕ from $\mathbf{n}$ [22]. Referring to Fig. 12, the cast shadow contour C for a potential target in the sonar image can be identified by either edge detection or computing the boundary of the shadow



(a) 1 (b) 2 (c) 3 (d) 4

(e) 5 (f) 6 (g) 7 (h) 8

Fig. 11. Regions around each object taken from different sonar images in the data set, showing that an object's shadow provides cues about object's shape and its occluding contour.

region. Taking each sea floor point **P** with spherical coordinates $\{\Re, \gamma, \phi\}$ on the cast shadow, the corresponding iso-azimuth frontal acoustic beam a can be traced to the occluding contour point $\mathbf{P}'$ with the spherical coordinates $\{\Re', \gamma, \phi\}$; this lies on the object/shadow boundary, and has a smaller range value $\Re' < \Re$ than the shadow/background boundary point **P**. Repeating for each cast shadow boundary point, this procedure allows us to reconstruct the entire 3-D occluding contour O of the object.

To summarize, we need only the positions of a minimum of 3 seabed features from the optical image, and their correspondences in the sonar images. Either distinct natural sea floor features or markers dropped on the seabed may be employed. The balance of information to construct each target's occluding contour in 3-D comes from the sonar image, by identifying the boundary of the shadow that is cast on the sea floor,

4) MAFR-TM Technique: The Multiple-Aspect Fixed-Range Template Matching (MAFR-TM) is a sonar-specific methodology to detect and classify the UXOs using the low resolution sonar

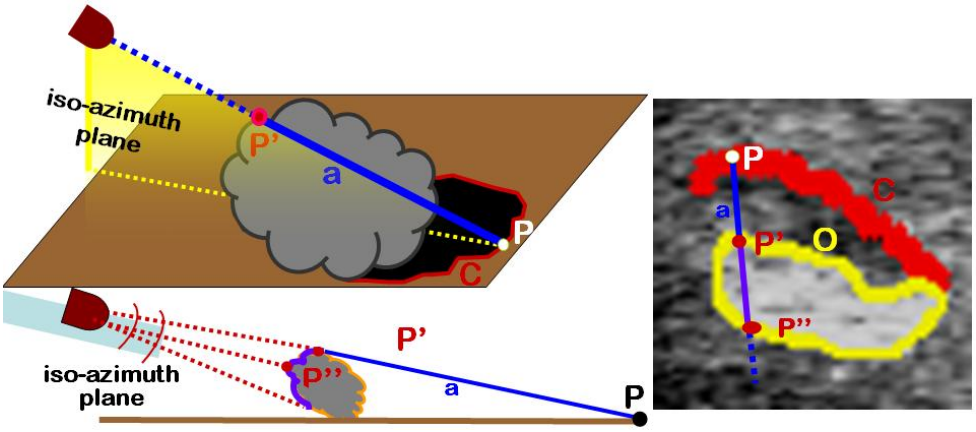


Fig. 12. Reconstruction of occluding contour from shadow boundary of sonar image C, by tracing from boundary points P along iso-azimuth frontal acoustic beams a to points P' on the occlusion boundary O.

data captured by the DIDSON US300. (It can be easily extended to the BlueView sonar data.) This technique is designed to detect and classify a target of high characteristic impedance in an environment that contains similar shaped objects of low characteristic impedance. The MAFR-TM algorithm is based on the proven concept of the template matching technique, which is a two-dimensional correlation between a reference image (template) and an image collected during field operations (source image). In the MAFR-TM algorithm, the template matching method is efficiently implemented in the wave number domain using two-dimensional Fast Fourier Transforms (2D-FFT) and wave number leakage is reduced with an optimized separable two-dimensional Kaiser window. MAFR-TM relies on the fact that UXO detection and classification is a re-acquisition process: the sonar is expected to scan a suspicious area, where UXO is likely to be present. Hovering vehicles (remotely piloted or autonomous) equipped with sector scan-sonar can follow a circular trajectory and create a 360 [deg] view of a small area (on the order of 20 [sqm]). MAFR-TM relies on a known vehicle trajectory to generate the templates under ideal conditions: in a tank with no clutter, limited bottom roughness and in the presence of a single target at a known aspect angle and range. The templates are grouped into classes, each class representing a specific object (in this case, a cylinder) observed at various aspect angles. Therefore, MAFR-TM does not need the two-tiered training-and-classification process of traditional computer-aided-detection-and-classification software. Instead, it can use a known database and perform the detection and classification "on the fly," so long as the vehicle keeps the potential target within a certain distance window. The platform does not have to follow exactly the same trajectory for MAFR-TM to work, as this technique defines areas where the target is likely to be present. The MAFR-TM implementation is covered in detail in Appendix A.

C. Multi-Modal DS-Based Data Fusion

The enormous difficulties associated with effective detection of the UXO, generally in harsh clutter-filled underwater environments, are related to several factors:

1) The rampant imperfections in the gathered data (missing, incomplete, uncertain and/or ambiguous entries).
2) The need for heterogeneous data of different modalities and resolutions to enhance discrimination capability. In fact, no single sensing technology has been both accurate and cost-effective for estimating even the density and spatial distribution of UXO objects, let alone identifying them. So, at any instant, all available data have to be exploited to the fullest extent possible. This may include the need to account for auxiliary evidence sources (e.g., domain experts) that often play critical roles despite their significant qualitative component (e.g., subjectivity of, and conflicts in, expert opinions).
3) Data sources may possess questionable reliability and integrity.
4) The need to utilize both stored (e.g., historical databases, domain expertise) and streaming (e.g., sensor) data.

Thus, the need of a framework capable of modeling and reasoning with data imperfections is obvious. Conventional approaches for effective detection of signatures encoded in the data typically utilize methods and tools from signal processing, Bayesian inference/learning, and of late, data mining and machine learning that are mainly posited on the Bayesian framework. To realize high detection accuracy in underwater environments, these methods must effectively deal with the various types of imperfections in both gathered evidence/data and signatures, viz., missing, incomplete, uncertain and/or ambiguous data, data plagued by severe clutter, multi-modal and heterogeneous data streams and stored data, qualitative aspects of domain expert information, UXO object signatures developed from imperfect evidence, etc.

When full probabilistic information is available, these Bayesian-based methods provide perhaps the best possible results. But, in the environment of concern here where data imperfections are rampant, it is unrealistic to assume the availability of prior information regarding the underlying probability distributions. Ignoring these imperfections and making unrealistic assumptions regarding the underlying probabilistic information can lead to inferences that are untrustworthy and detection performance that is unacceptable.

What we need is a framework that is capable of properly modeling, and not neglecting, the various data imperfections throughout the process from raw data to fused information to decision-making. The Bayesian framework is not well suited for this task. Its effectiveness is muted by the requirement of numerous a-priori assumptions about priors (e.g., equi-probability allocation when a value is missing assumes that it is missing-at-random (MAR)) and models that are difficult to justify and ill-suited for environments rife with data imperfections. The main theoretical theme of this project was the assessment of a novel data fusion paradigm to enable decision making based on imperfect heterogeneous data.

This data fusion paradigm is based on the *Dempster-Shafer (DS) belief theoretic framework* [32] which appears to possess certain advantages over other alternatives for representing and

working with data imperfections [2]–[4], [37]. During the last few decades, DS theory has emerged as one of the most dominant frameworks in uncertainty processing. This recent interest in DS theory is evident from the wide spectrum of problem domains it has already been applied to. These applications range from, but not limited to, basic signal processing techniques, such as, detection [6], [8], [9] and identification [5], [10], [17], to general information filtering techniques, such as, rule mining [14], collaborative filtering [34], and others [7], [11], [13]. The recent interest in fusing the more *qualitative* 'soft' information [36] into the decision making process has resulted in more flexible DS theoretic evidence updating and combination strategies [33].

In the work described below, we explore how these latest DS theoretic evidence updating schemes can be used to improve the underwater UXO detection and characterization tasks. The use of these updating schemes enables one to:

1) work with heterogeneous data possessing non-identical 'scopes of expertise',
2) conveniently account for source reliability, and
3) make use of 'qualitative' domain-expert evidence.

As stated, we had scheduled for the application to existing magnetometer and electromagnetic array data from ESTCP, as well as to 2-D optical and sonar data collected in an outdoor tank and in the ocean. Unfortunately, this data lacked ground truth to establish quantitative assessment, and so the recorded optical and sonar imagery were the basis for illustrating these techniques via an experiment to characterize underwater objects using optical and acoustic cameras. The detection 'signatures' of objects are chosen to be the acoustic shadow characteristics.

V. Dempster-Shafer (DS) Theory Related Work

A. Basic Notions

In DS theory, the total set of mutually exclusive and exhaustive propositions of interest (i.e., the 'scope of expertise') is referred to as the *frame of discernment (FoD)* $\Theta = \{\theta_1, \ldots, \theta_n\}$ [32]. A singleton proposition θ_i represents the lowest level of discernible information. Elements in the power set of Θ, 2^Θ, form all the propositions of interest. We use $A \setminus B$ to denote all singletons in A that are not in B; $\overline{A}$ denotes $\Theta \setminus A$.

Definition 1. *Consider the FoD* Θ *and* $A \subseteq \Theta$.

1) *The mapping* $m_\Theta(\bullet) : 2^\Theta \mapsto [0,1]$ *is a* basic belief assignment (BBA) *or* mass assignment *if* $m_\Theta(\emptyset) = 0$ *and* $\sum_{A \subseteq \Theta} m_\Theta(A) = 1$. *The BBA is said to be* vacuous *if the only proposition receiving a non-zero mass is* Θ.
2) *The* belief *of* A *is* $Bl_\Theta(A) = \sum_{B \subseteq A} m_\Theta(B)$.
3) *The* plausibility *of* A *is* $Pl_\Theta(A) = 1 - Bl_\Theta(\overline{A})$.

■

The DS theory models the notion of *ignorance* by allowing the mass to composite propositions (i.e., a non-singleton proposition). A proposition that possesses non-zero mass is a *focal element.* The set of focal elements is the *core* $\mathfrak{F}_\Theta$*;* the triple $\{\Theta, \mathfrak{F}_\Theta, m_\Theta(\bullet)\}$ is the corresponding *body of evidence (BoE).* While $m_\Theta(A)$ measures the support assigned to proposition A *only,* the belief

represents the total support that can move into A without any ambiguity; $Pl_\Theta(A)$ represents the extent to which one finds A plausible. When focal elements are constituted of singletons only, the BBA, belief and plausibility all reduce to a probability assignment.

B. Evidence Combination

Definition 2 (Dempster's Combination Rule (DCR))**.** *The DCR-fused BoE* $\mathcal{E} \equiv \mathcal{E}_1 \oplus \mathcal{E}_2 = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\bullet)\}$ *generated from the BoEs* $\mathcal{E}_i = \{\Theta_i, \mathfrak{F}_{\Theta_i}, m_{\Theta_i}(\bullet)\}$, $i = 1, 2$, *when* $\Theta \equiv \Theta_1 = \Theta_2$, *is*

$$m_\Theta(A) = \sum_{C \cap D = A} m_{\Theta_1}(C)\, m_{\Theta_2}(D)/(1-K),\ \forall A \subseteq \Theta,$$

whenever $K = \sum_{C \cap D = \emptyset} m_{\Theta_1}(C)\, m_{\Theta_2}(D) \neq 1$. ■

Note that $K \in [0, 1]$ is an indication of the conflict between the evidence provided by the BoEs. Hence, K is referred to as the *conflict* between the BoEs being fused. The DCR's difficulties in fusing conflicting BoEs are well documented. The requirement that the two FoDs being fused be identical constitutes another drawback associated with the DCR.

To fuse evidence generated from non-identical FoDs Θ_1 and Θ_2 (so that $\Theta_1 \neq \Theta_2$ and $\Theta_1 \cap \Theta_2 \neq \emptyset$), one can simply ignore the differences in the FoDs by having each source allocate zero mass to propositions that are not within its own FoD and continue applying DCR. So, this approach assumes that each source can discern $\Theta_1 \cup \Theta_2$ and ignores the fact that some propositions are not within its scope of expertise. The counter-intuitive conclusions this approach may generate are well documented [15]. In the deconditioning approaches, each source would artificially introduce ambiguities into its evidence so that its own FoD is 'expanded' to $\Theta_1 \cup \Theta_2$.

C. Conditional Update Equation (CUE)

The conditional approach to fusing evidence 'conditions' or 'updates' the already available evidence with respect to what both FoDs can discern [16]. Once the conditioning operation is performed, each source invokes a strategy to incorporate its originally cast evidence that does not belong to $\Theta_1 \cap \Theta_2$. This approach enables a source to update its own knowledge base, and exchange information with other sources for the express purpose of refining its own knowledge, without having to continually 'expand' its FoD.

The *conditional update equation (CUE)* in [33] embraces this conditional approach. The identical FoDs case of CUE in [26] is applicable when FoDs of all evidence sources to be fused are identical. To explain, consider the two BoEs $\mathcal{E}_i[k]$ with $\Theta \equiv \Theta_1 = \Theta_2$. The CUE in [26] then yields the update $\mathcal{E}_1[k+1] \equiv \mathcal{E}_1[k] \lhd \mathcal{E}_2[k]$, $\forall k \geq 0$, of $\mathcal{E}_1[k]$ as

$$Bl_{\Theta_1}(B)[k+1] = \alpha[k]\, Bl_{\Theta_1}(B)[k] + \sum_{A \subseteq \Theta_2} \beta(A)[k]\, Bl_{\Theta_2}(B|A)[k],$$

where $Bl_{\Theta_2}(A) > 0$ and the parameters $\{\alpha[\bullet], \beta(A)[\bullet]\}$ are non-negative and satisfy

$$\alpha[k] + \sum_{A \subseteq \Theta_2} \beta(A)[k] = 1, \forall k \geq 0, \tag{2}$$

with $\beta(A)[\bullet] = 0$, $\forall A \notin \mathfrak{F}_{\Theta_2}[\bullet]$.

The equivalent mass update equation has the same form and is given by

$$m_{\Theta_1}(B)[k+1] = \alpha[k]\, m_{\Theta_1}(B)[k] + \sum_{A \subseteq \Theta_2} \beta(A)[k]\, m_{\Theta_2}(B|A)[k].$$

The conditional operation in the CUE is implemented using the Fagin-Halpern (FH) DS theoretic conditionals [12], as described below:

Definition 3. *For $\mathcal{E} = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\bullet)\}$, $A, B \subseteq \Theta$ with $Bl_\Theta(A) > 0$, the* conditional belief *of B given A is*

$$Bl_\Theta(B|A) = \frac{Bl_\Theta(A \cap B)}{Bl_\Theta(A \cap B) + Pl_\Theta(A \setminus B)}.$$

The conditional plausibility *$Pl_\Theta(B|A)$ is given by a counterpart equation.* ■

A result that can be used to efficiently implement the CUE has recently appeared in [35]:

Theorem 1 (Conditional Core Theorem (CCT))**.** *Consider the BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\bullet)\}$. Then, $m(B|A) > 0$, $A \in \hat{\mathfrak{F}}$, iff B can be expressed as $B = X \cup Y$, for some $X \in in(A)$, $Y \in OUT(A) \cup \{\emptyset\}$, where $in(A) = \{B \subseteq A \mid B \in \mathfrak{F}\}$ and $out(A) = \{B \subseteq A \mid B \cup C \in \mathfrak{F}, \emptyset \neq B, \emptyset \neq C \subseteq \overline{A}\}$.* ■

Remark: The conditional computations may take significant amount of time and computational resources, especially when the FoDs are large. The CCT can be used to efficiently implement the CUE and other conditional based algorithms.

D. DS Theoretic Characterization of Optical and Acoustic Data and Problem Formation

Optical cameras are very useful for identifying the objects in relatively clear water, especially when the objects have sufficient texture or surface markings. However, they can quickly become ineffective is turbid waters. However, sonar cameras are robust under these conditions since acoustic energy can *penetrate through silt*, although identifying objects from sonar images is not so trivial. In addition, there can be significant variations in acoustic signatures with type/size of munitions, whether these are fully intact, distorted or broken into munitions-related scrap, filled/empty, buried, partially buried or proud. We next explore how the DS theoretic framework may be applied to fuse the visual cues in optical and sonar images, overcoming their shortcomings and exploiting their advantages.

If the positioning of the acoustic camera with respect to the target is known, and if the shape and dimensions of the sought after target is known, the shadow of the target can be characterized under certain conditions. For instance, spherical and cylindrical targets cast different shadows. The discrimination may be difficult when the targets are laid on the ground. However, this can be overcome by analyzing the shadow from different perspectives or viewing angles. One can also make use of the evidence from optical cameras to improve the characterization task if/when such information is available. To reduce the uncertainty and for robust decision making, evidence from both optical and acoustic cameras over different views can be combined.

Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ be the set of targets of interest. Once an object is observed, optical sensor provides the BoE $\mathcal{E}_O = \{\Theta, \mathfrak{F}_O, m_O(\bullet)\}$ which models the optical evidence. Similarly, sonar sensor provides the BoE $\mathcal{E}_S = \{\Theta, \mathfrak{F}_S, m_S(\bullet)\}$. The objective is to obtain the fused BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\bullet)\}$ to characterize the object observed. We seek for a method to obtain the fused BoE $\mathcal{E}$ for the task at hand by exploring different possible ways of fusing the available evidence.

Suppose an object, say θ, is viewed. We want to determine if this object θ is the same as one of the target types $\theta_1, \ldots, \theta_n$. For the viewed object, we come up with two models for optical and sonar camera views. For instance, take the DS theoretic model $\mathcal{E}_O$ for the evidence obtained from the optical camera. This model represents how likely θ belongs to one of the target types mentioned above based on the evidence obtained from the optical camera. Say we have $m(\theta_1) = 0.9$, $m(\theta_2) = 0.1$, for a certain view. This can be interpreted as evidence in support of the viewed object being θ_1 with 90% confidence, and θ_2 with 10% confidence. Now, we form models (i.e., BoEs) for both optical (BoE is $\mathcal{E}_O$) and sonar (BoE is $\mathcal{E}_S$) for each view, i.e., for viewing angles $\alpha = 0, 30, \ldots, 150$. The objective is to explore different ways of fusing this evidence (i.e. combining available BoEs) to characterize θ into one of the target types $\theta_1, \ldots, \theta_n$.

VI. Results

A. Test Site Mosaic Generation

As stated, optical images were not acquired in the marina due to the high turbidity during our scheduled data acquisition timetable. As a result, the construction of the site photo-mosaic was limited to the data collected in the FAU outdoor pool data. Collecting stereo data, we have employed the images from the left and right cameras, allowing us to minimize the number of adjacent swaths for data collection.

Fig. 13 depicts the camera platform trajectory, positions where images were recorded, the mosaic, as well as samples of the raw data prior to pre-processing for the application of the algorithm. The trajectory and the mosaic was generated from a total of 390 out of roughly 1500 original frames, chosen automatically by our photo-mosaicing system over 6 horizontal and 6 vertical swaths (3 for each of left and right cameras) covering a 4.5 [m] $\times$4.5 [m] area.

As one can readily verify, the bottom surface has relatively weak texture and the visual cues for frame-to-frame registration is limited to regions with prominent features, e.g., the edges at the object boundaries. The dynamic range and color adjustment, which is readily noted by comparing the mosaic with the raw frames, is offset by the impact of background noise enhancement. Despite misalignment in some local regions, the mosaic quality is sufficient to obtain a relatively accurate estimate of the spatial arrangement of various targets within the scene. Where necessary to improve the quality for the same terrain texture content, image density and overlap can be increased by reducing the distance between adjacent swaths.

When operating near the sea floor, the the natural terrain objects typically have stronger texture, so the alignment can be further improved. To appreciate the improvement in alignment for a scene with strong texture, the mosaic for a short segment of data collected in the Bahamas in an earlier SERDP project is depicted in Fig. 14 [27]. As noted from the camera trajectory, the coverage is an area of 1 [m] $\times$ 8 [m], with spatial resolution of roughly 1 [mm].

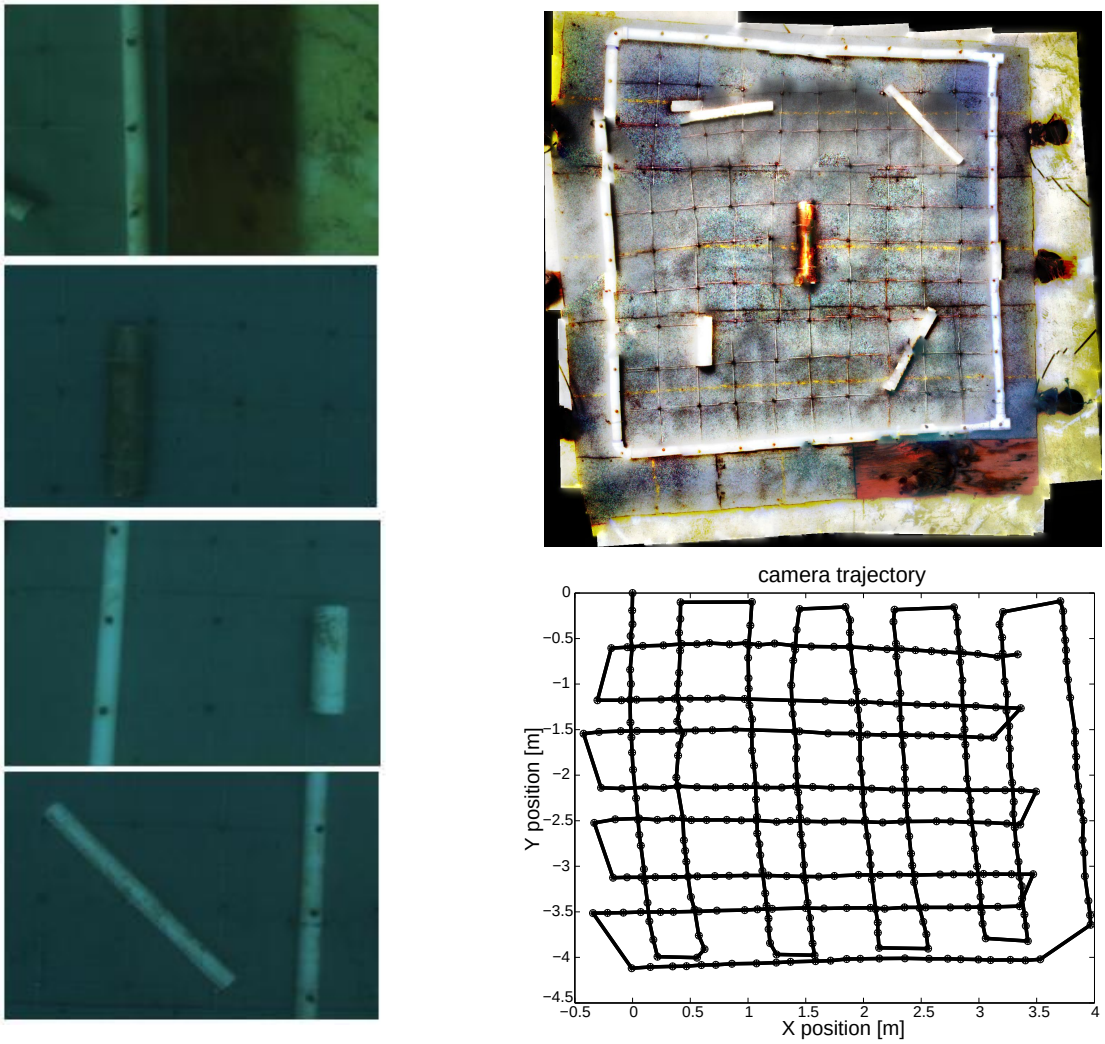


Fig. 13. Sample raw images used for construction of site mosaic and computer camera trajectory.

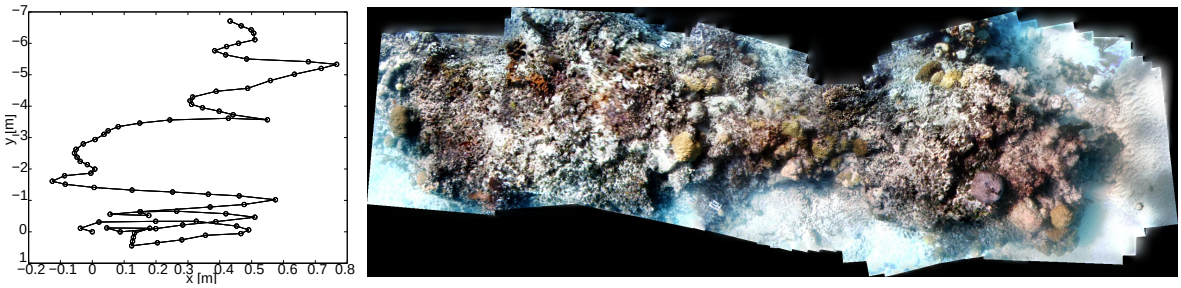
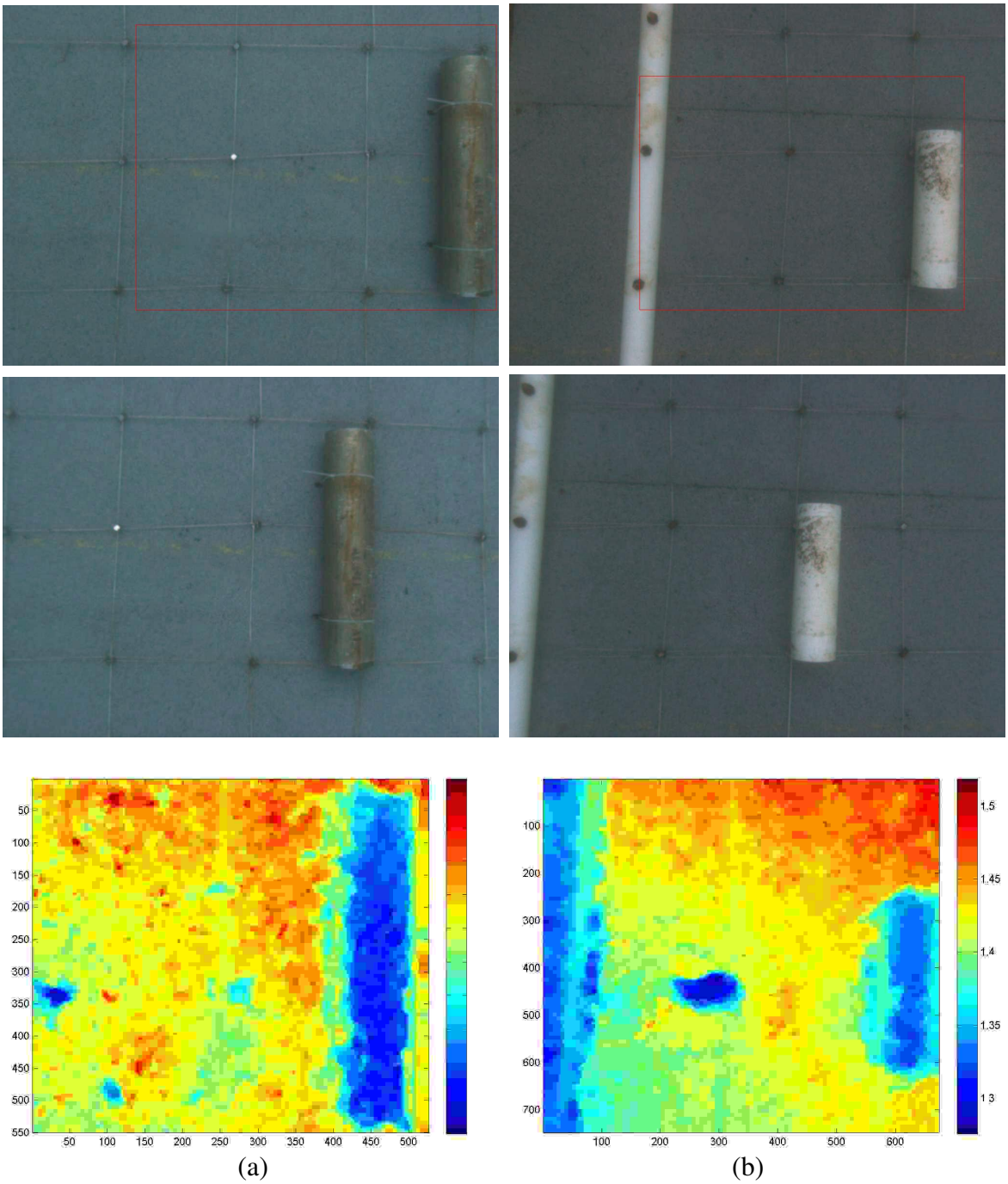


Fig. 14. Camera trajectory for sample mosaic of natural terrain constructed from data in Andros Island, Bahamas.

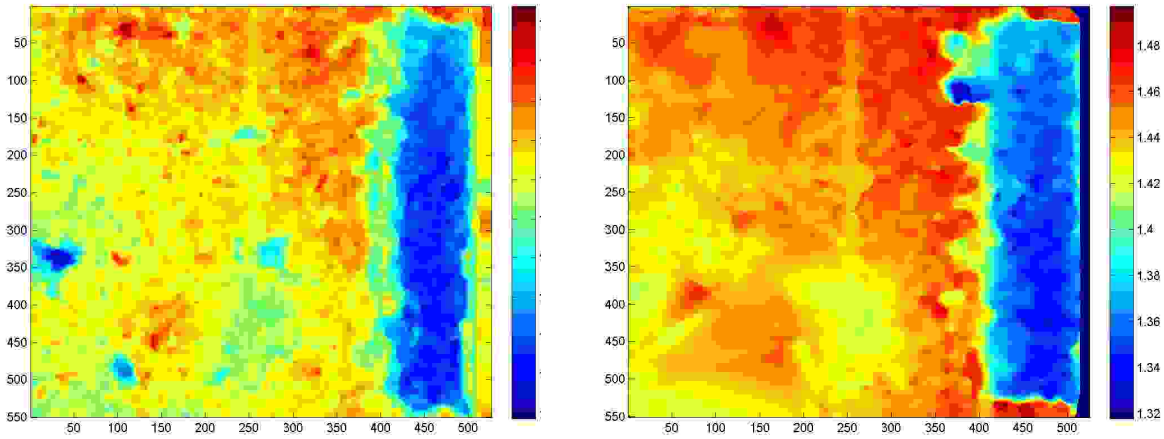


(a) (b)

Fig. 15. Stereo pairs and 3-D reconstruction of (a) steel target, and (b) PVC pipes. Distance from scene shown by color coding.

B. 3-D Reconstruction from Stereo Data

As explained earlier, it is sufficient to detect an object in any view of sensing modality in order track it in other overlapping views and modalities. When the visibility allows to acquire



(a) (b)

Fig. 16. Application of two different stereo techniques for 3-D scene reconstruction.

data with sufficient details, the stereo data corresponding to a sensor platform position where the target of interest is viewed can be processed for the construction of a 3-D topographical map. Using the pool data, we demonstrate this for two particular stereo pairs, one for the steel cylinder and the other for two PVC pipes; see left and right views in the top two rows of Fig. 15.

For the two cases, we have shown the color-coded depth maps – distance from the cameras to the scene and its objects – for the region highlighted by the red boxes in the left views. We have applied a dense multi-resolution optical flow-based algorithm [20]. Some observations are in order: 1) both the steel and PVC pipes are sized correctly at 10 [cm]; 2) there is a slight bottom surface tilt from top-right to bottom-left; 3) the reconstruction is more noisy in (b), which is directly tied to the weaker texture as evident in the stereo pairs; e.g., note the noisy blue blob roughly in the center of the 3-D map.

In Fig. 16, we have compared the results in Fig. 15(a) with the estimate from the application of another algorithm. In this other method, we start with certain matched prominent features as seeds, and then a dense reconstruction is obtained by a fast efficient local match propagation strategy [18]. While the results are consistent over the target region, this second method gives a noisier estimate over the pool bottom surface, mainly because the propagation is somewhat ambiguous over these regions with weaker texture. In fact, the algorithm is not unable to propagate many of the seed matches for the PVC pipes scene, thus, the reconstruction is very sparse; not given here.

C. Registration and 3-D Reconstruction of Occluding Contours

As we explained we can applying the opti-acoustic stereo solution to determine the 3-D positions $\mathbf{P}$ of a minimum of 3 seabed matching features $\{\mathbf{p}_s, \mathbf{p}_o\}$. These allow us to fix the plane $\mathbf{n}$ of the sea floor. With the Knowledge of the plane, we can determine the 3-D position

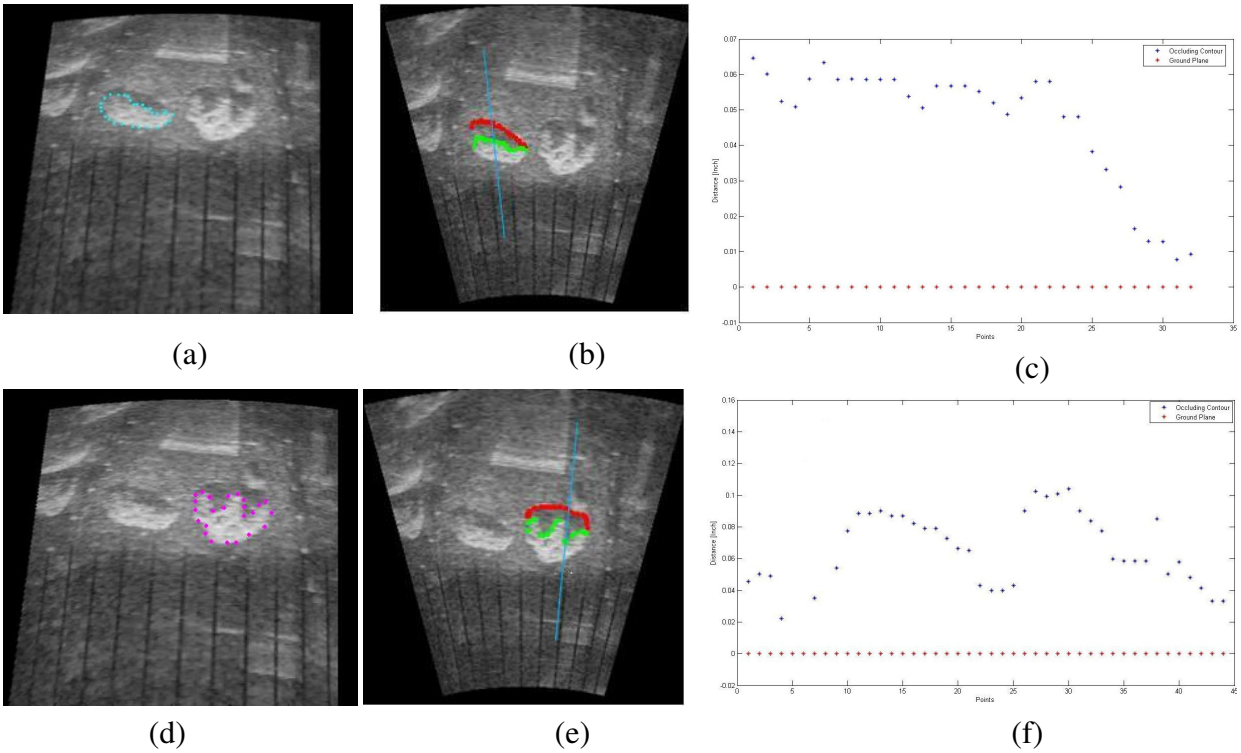


Fig. 17. (a,d) Two coral objects, (b,e) occluding boundaries in sonar image, and (c,f) 3-D reconstruction of these boundaries, relative to the flat bottom surface, represented by red lines. Object size can be deduced from the contour reconstruction.

of any sea floor feature in the sonar image, particularly the boundary points on the object's cast shadow. Tracing the iso-azimuth frontal acoustic beam to the occlusion contour allows us to determine the countour in 3-D.

Referring back to the data in Figs. 9(a), shown again in Fig.17, consider the two corals with non-uniform contour. As demonstrated in (b) and (e), the tracing of shadow contours to the occluding contours leads to the reconstruction depicted in (c) and (f), where the bottom surface of the flat background surface has been represented by the red lines. While the accuracy of the reconstruction depends on the ability to detect the shadow and occluding contours, it can often suffice to discriminate between natural objects with arbitrary contour shapes and manmade objects. Also, the height relative to the bottom surface provides an estimate of object size. While not utilized here, it is noted that the object's lower boundary in the sonar view does not represent a visibility rim, but rather points $\mathbf{P}''$ with smallest range for different azimuth directions, as depicted in Fig. 12.

D. MAFR-TM Technique

Tables 1-11 summarize the results from the application of the MAFR-TM algorithm to detect, establish the orientation and location of, and classify the target in the low-resolution sonar images

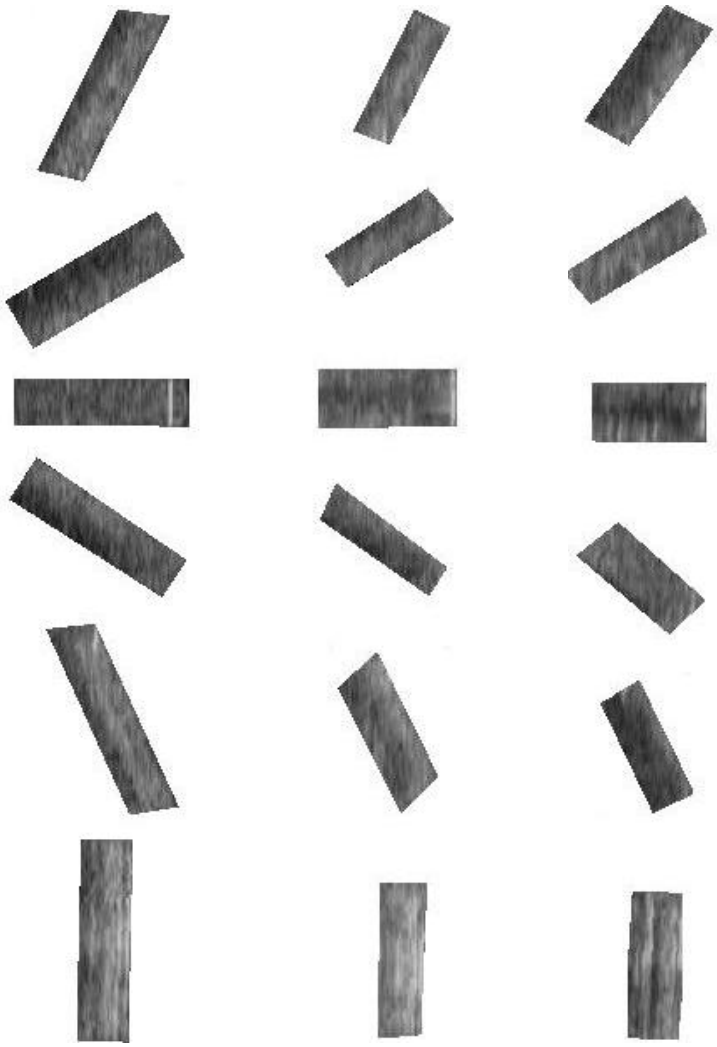


Fig. 18. Templates of steel (left), aluminum (middle), and PVC (right) cylindrical targets at 6 orientations (for 0, 30, 60, 90, 120, and 150 [deg]) relative to the sonar.

captured by the DIDSON US300. The data sets used for this study are referred to as 'Single Target Tank Data," 'Cluttered Tank Data," and 'Cluttered Marina Data."

The algorithm has performed fairly well with the tank and cluttered tank data, particularly for the target detection, and the discrimination between metals and plastics. There is room for improvement in determining the orientation and location of the detected target.

The steel target has been detected 100% of the time in singular tank data, 94% of the time in a cluttered tank data set, and 61% in cluttered marina data. The majority of the undetected targets occurred at 90 [deg] and 270 [deg] orientations due to the very small echo the sonar perceives.

Fig. 18 depicts examples of the templates (for 0, 30, 60, 90, 120, and 150 [deg]) for each

Tank Data			Marina Data		
Source Image Angle	*Detected? (Before)*	*Detected? (After)*	*Source Image Angle*	*Detected? (Before)*	*Detected? (After)*
0˚	Yes	Yes	0˚	Yes	Yes
10˚	Yes	Yes	10˚	Yes	Yes
20˚	Yes	Yes	20˚	Yes	Yes
30˚	Yes	Yes	30˚	Yes	Yes
40˚	Yes	Yes	40˚	Yes	Yes
50˚	Yes	Yes	**50˚**	**No**	Yes
60˚	Yes	Yes	**60˚**	**No**	Yes
70˚	Yes	Yes	70˚	Yes	Yes
80˚	Yes	Yes	80˚	Yes	Yes
90˚	Yes	Yes	90˚	Yes	Yes
100˚	Yes	Yes	100˚	Yes	Yes
110˚	Yes	Yes	110˚	Yes	Yes
120˚	Yes	Yes	120˚	Yes	Yes
130˚	Yes	Yes	**130˚**	**No**	Yes
140˚	Yes	Yes	**140˚**	**No**	Yes
150˚	Yes	Yes	**150˚**	**No**	Yes
160˚	Yes	Yes	160˚	Yes	Yes
170˚	Yes	Yes	**170˚**	**No**	Yes
180˚	Yes	Yes	**180˚**	**No**	Yes

TABLE I
TARGET DETECTION RESULTS WITH STEEL OBJECT BEFORE AND AFTER CTR PREPROCESSING.

of the three targets, used in the classification part of the methodology, as well as determining the orientation and location of the labeled target. The algorithm has classified a steel target 67% of the time in singular tank data. This relatively poor performance is because the algorithm sometimes cannot discriminate metal from plastic *in images of singular targets with weak echos*. The MAFR-TM approach in the wave number domain is a pixel multiplication process. Steel generates the strongest echo of the three template classes, and when this template class is matched (or multiplied) with a weak echo or background intensity, the algorithm may compute a higher correlation and in turn a larger SNR than other template classes. Therefore the algorithm may confuse a singular PVC target with that of a singular steel target in singular object scenes. In the case of multiple object scenes, such as in the cluttered tank and marina data, the classification success rate is no less than 75%. Here, the algorithm would not distinguish well between metal classes, but could discriminate between metal and plastic classes. The confusion between the two metal classes is caused by their rather similar returns or intensities to the sonar.

Data Type	*Detected*	*Total Images*	*Detection Rate*
Tank (Before)	19	19	100.0%
Tank (After)	19	19	100.0%
Marina (Before)	12	19	63.2%
Marina (After)	19	19	100.0%

TABLE II
SUMMARY OF DETECTION RESULTS FOR STEEL TARGET, BEFORE AND AFTER CTR PREPROCESSING.

In addition, the orientation of a detected object could often be determined with reasonable accuracy. As Table VII illustrates, the algorithm achieved a minimum success rate of 57.9% over all four cases. It also yielded the position of a detected object within roughly 90% of its true location, when the target type and orientation is established correctly (i.e., within 60 [pix] in images with lowest resolution of 575 [pix] $\times$ 568 [pix]). The majority of errors in the singular object tank and cluttered tank data occur at (or close to) aspect angles of 90 [deg] and 270 [deg]. Again, spurious errors in between may have been caused by wall reflections present within the source image.

The performance decreased by approximately 30% when applied to the marina data set. In this case, many factors influence the outcome of the technique, including the partial/complete burial in soft mud, in addition to clutter. When the target is partially buried or obstructed by clutter, the template cannot completely overlap the true target in the source image. The partial alignment can lead a lower correlation with the template, relative to an unobstructed false target or a bright reflection from the seafloor. Also, the targets return can be very weak (sometimes resembling the intensity of a shadow) at certain aspect angles, due to cast shadow of a fish, clutter, or other obstructions on the seafloor. As the sonar is rotated around the target to perform the 360 [deg] survey, the targets response alternates from very strong (bright) to very weak (dark). This is why it is important to capture as many views of the suspected UXO area from different angles, to overcome the ambiguities of *"degenerate views."*

Finally, the MAFR-TM technique has also been applied to pre-processed data, comprising the steel cylinder as the true target. The first set is the same singular object scene collected in tank, and the second is the same cluttered scene acquired in the marina. The preprocessing stage consisted of applying a crosstalk reduction (CTR) algorithm to each source image in the database. The CTR technique generally sharpens the image by eliminating the averaging effect due to the crosstalk between adjacent beams. Thus, strong returns become brighter and weaker returns darker, as depicted in Figs.19.

The CTR processing was carried out at the very end of our project (early October'10), based on an ongoing product development work at Sound Metrics Corp.[5]. This was arranged through

[5]Manufacturer of DIDSON systems.

Tank Data			Marina Data		
Source Image Angle	*Class? (Before)*	*Class? (After)*	*Source Image Angle*	*Class? (Before)*	*Class? (After)*
0˚	Steel	Aluminum	0˚	PVC	Aluminum
10˚	Steel	Aluminum	10˚	PVC	PVC
20˚	Steel	Aluminum	20˚	Aluminum	Aluminum
30˚	Steel	Aluminum	30˚	Aluminum	Aluminum
40˚	Aluminum	Aluminum	40˚	Aluminum	Aluminum
50˚	Aluminum	Aluminum	50˚	-	Aluminum
60˚	Steel	Aluminum	60˚	-	Aluminum
70˚	Aluminum	Aluminum	70˚	Aluminum	Aluminum
80˚	Steel	Aluminum	80˚	Aluminum	Aluminum
90˚	Steel	Aluminum	90˚	PVC	Aluminum
100˚	Steel	Aluminum	100˚	Aluminum	Aluminum
110˚	Steel	Aluminum	110˚	Aluminum	Aluminum
120˚	Steel	Aluminum	120˚	-	Aluminum
130˚	Aluminum	PVC	130˚	-	Aluminum
140˚	Steel	Aluminum	140˚	-	Aluminum
150˚	Aluminum	Aluminum	150˚	Aluminum	Aluminum
160˚	Steel	PVC	160˚	-	Aluminum
170˚	Steel	Aluminum	170˚	-	Aluminum
180˚	Steel	Aluminum	180˚		

TABLE III
CLASSIFICATION RESULTS FOR A STEEL TARGET BEFORE AND AFTER CTR PREPROCESSING.

some discussions about our SERDP project during the IEEE/MTS Oceans'10 Conference (held in Seattle, WA, September 21-23, 2010) with the company's technical staff. Subsequently, our results based on limited experiments are only preliminary, and the advantage of CTR preprocessing stage at this stage is only circumstantial. Further research and extensive experiments are warranted to establish a true performance assessment.

Since the targets are cylindrical, and thus symmetrical about all axes, only the data for 0 [deg] - 180 [deg] were used to explore the impact of CTR processing on the MAFR-TM Algorithm in the detection, classification, orientation, and location of a steel target. The results have been obtained for four data sets, both the single target tank data and the cluttered marina data sets. For the tank data, there is no improvement in classification performance, and often degradation, with the CTR data. This can be readily explained by noting that our templates of the three objects were constructed from unprocessed images acquired under perfect tank conditions. Thus, it is expected that when testing with same data type, they correlate better with object returns in the

Data Type	*Classified*	*Percentage*	*Classification Success Rate*	
Tank (Before)				
Steel	14	73.7%	Metal	100%
Aluminum	5	26.3%	Plastic	0.0%
PVC	0	0.0%	MP Ratio	1:0
Total Detected	19	-		
Tank (After)				
Steel	0	0%	Metal	100%
Aluminum	17	89.5%	Plastic	0.0%
PVC	2	10.5%	MP Ratio	1:0
Total Detected	19	-		
Marina (Before)				
Steel	0	0.0%	Metal	75.0%
Aluminum	9	75.0%	Plastic	25.0%
PVC	3	25.0%	MP Ratio	3:1
Total Detected	12	-		
Marina (After)				
Steel	0	0.0%	Metal	89.5%
Aluminum	17	89.5%	Plastic	10.5%
PVC	2	10.5%	MP Ratio	8.5:1
Total Detected	19	-		

TABLE IV
SUMMARY OF CLASSIFICATION RESULTS FOR STEEL TARGET, BEFORE AND AFTER CTR PREPROCESSING.

unprocessed data, compared to CTR processed data.

For marina data, the CTR algorithm enhances the MAFR-TM's detection somewhat, mainly because the improved dynamic range enhances the detection of cast shadow regions as well as target returns. The same reason for the pool data results applies in explaining the classification performance, particularly in discriminating between metal classes. However, the algorithm can classify between metals and plastics. Unfortunately, the inability to differentiate between steel and aluminum directly and adversely affects the estimating of the target orientation and location.

1) Target Detection: The detection results for data with targets at various orientations are illustrated in Table I, and a summary of the detection performance in TableII. We note that the MAFR-TM algorithm performs rather well when applied to the tank data. Before pre-processing, the performance decreases by about 30% for the marina data, mainly due to significant intensity variations in target and shadow zones, because the scene contains not only the true target, but significant clutter and false targets; see Figs. 19(b,b'), which is enhanced in the CTR processed

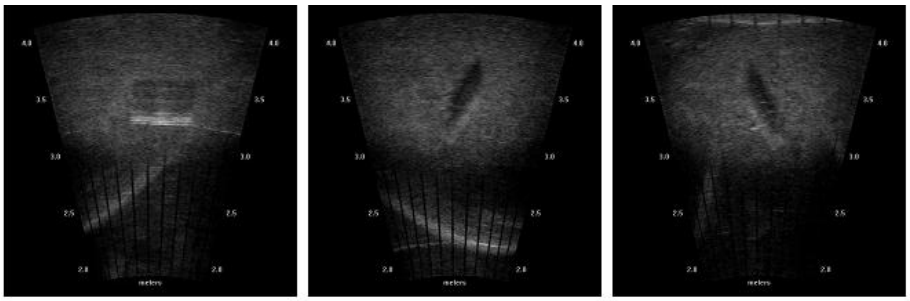
Tank Data				
Source Image Angle	*Orientation? (Before)*	*Orientation? (After)*	*Angle Error (Before)*	*Angle Error (After)*
0°	0°	20°	0°	20°
10°	10°	50°	0°	40°
20°	20°	60°	0°	40°
30°	30°	90°	0°	60°
40°	40°	**180°**	0°	**140°**
50°	40°	20°	10°	30°
60°	50°	10°	10°	50°
70°	**0°**	50°	**70°**	20°
80°	30°	50°	50°	30°
90°	120°	60°	30°	30°
100°	110°	50°	10°	50°
110°	120°	10°	10°	100°
120°	130°	150°	10°	30°
130°	130°	160°	0°	30°
140°	140°	170°	0°	30°
150°	140°	170°	10°	20°
160°	160°	160°	0°	0°
170°	170°	10°	0°	20°
180°	180°	10°	0°	10°

TABLE V
ESTIMATING TARGET ORIENTATION, BEFORE AND AFTER CTR PREPROCESSING, FOR TANK DATA.

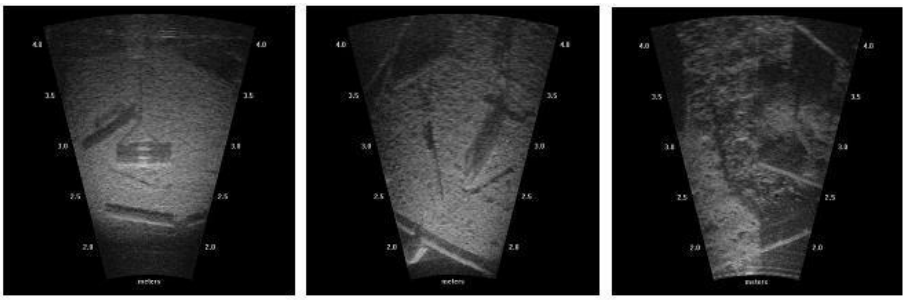
data. Noting that the shadow region is a strong visual cue for the presence of a 3-D object, the algorithm sometimes favors a shadow zone from clutter or a false target instead of the cast shadow of the true target, thus estimating a large distance between each target and its shadow zone. This leaves the target undetected in certain orientations. After data preprocessing, both the target and its shadow are enhanced, leading to correct detection.

2) Classification: Using the SNR, the proper target class (steel, aluminum, and PVC) can be identified by comparing the maximum correlation between the template classes over all template scans. A target is identified if the peak SNR corresponds to the same class across all (or a high percentage of) aspect angles. Tables III and IV depict the classification results of the tank and marina data sets, before and after preprocessing.

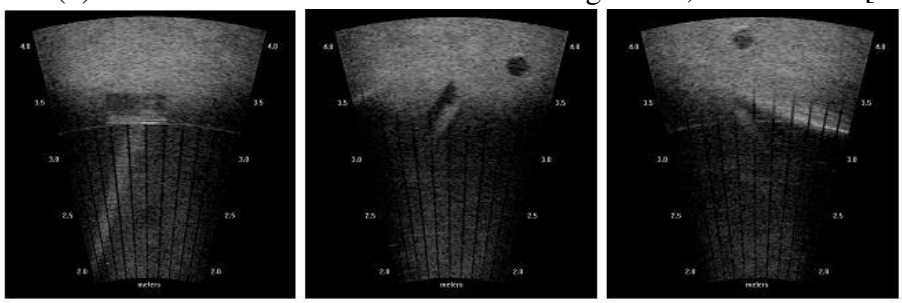
Before preprocessing, the best template class matching to the tank source images of the steel



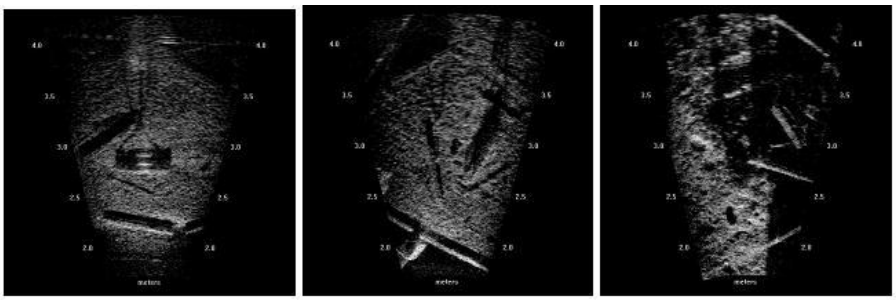
(a) Single-target pool images (180, 240, 300 [deg])



(b) Cluttered marina scene with steel target at 0, 60 and 120 [deg]



(a') image in (a) after CTR processing



(b') image in (b) after CTR processing

Fig. 19. Sample DIDSON images (a,b) before and (a',b') after application of CTR algorithm, from (a,a') pool and (b,b') marina data sets.

Marina Data				
Source Image Angle	*Orientation? (Before)*	*Orientation? (After)*	*Angle Error (Before)*	*Angle Error (After)*
0˚	20˚	20˚	20˚	20˚
10˚	40˚	20˚	30˚	20˚
20˚	30˚	20˚	10˚	0˚
30˚	80˚	10˚	50˚	20˚
40˚	10˚	10˚	30˚	30˚
50˚	-	140˚	-	90˚
60˚	-	150˚	-	90˚
70˚	**150˚**	50˚	**80˚**	20˚
80˚	140˚	50˚	60˚	30˚
90˚	130˚	120˚	40˚	30˚
100˚	110˚	160˚	10˚	60˚
110˚	110˚	120˚	0˚	10˚
120˚	120˚	20˚	0˚	100˚
130˚	-	**20˚**	-	**110˚**
140˚	-	50˚	-	90˚
150˚	-	50˚	-	100˚
160˚	140˚	110˚	20˚	50˚
170˚	-	150˚	-	20˚
180˚	-	190˚	-	10˚

TABLE VI
ESTIMATING TARGET ORIENTATION, BEFORE AND AFTER CTR PREPROCESSING, FOR MARINA DATA.

Data Type	*Detected*	*Min Angle Error*	*Max Angle Error*	*No. of Errors < or = 30*	*Success Rate*
Tank (Before)	19	0˚	70˚	17	89.5%
Tank (After)	19	0˚	140˚	12	63.1%
Marina (Before)	12	0˚	80˚	8	66.7%
Marina (After)	19	0˚	110˚	11	**57.9%**

TABLE VII
SUMMARY OF RESULTS IN ESTIMATING TARGET ORIENTATION, BEFORE AND AFTER CTR PREPROCESSING.

target at 19 orientations in 0 [deg] - 180 [deg] results in 14 correct classification as steel, 5 as aluminum, and none as PVC. This leads to the conclusion that the target is in fact a steel

Tank Data **Red: x-error > 60 pix, Blue: y-error > 60 pix, Green: x-error and y-error > 60 pix**					
Source Image Angle	*Actual (x,y) Position*	*Calculated (x,y) (Before)*	*Calculated (x,y) (After)*	*Error in (x,y) (Before)*	*Error in (x,y) (After)*
0˚	252, 197	249, 203	360, 82	3, 6	**108**, **115**
10˚	249, 200	245, 195	335, 75	4, 5	**86**, **125**
20˚	251, 193	255, 187	328, 67	4, 6	**77**, **129**
30˚	252, 178	256, 178	336, 62	4, 0	**84**, **116**
40˚	257, 167	260, 178	326, 106	3, 11	**69**, **61**
50˚	268, 163	264, 175	403, 68	4, 12	**135**, **95**
60˚	270, 156	256, 169	370, 62	14, 13	**100**, **94**
70˚	268, 171	224, 444	352, 54	44, **273**	**84**, **117**
80˚	274, 173	220, 414	324, 58	54, **241**	50, **115**
90˚	279, 177	241, 194	317, 52	38, 17	38, **125**
100˚	270, 177	257, 172	417, 53	13, 5	**147**, **124**
110˚	265, 181	254, 195	403, 75	11, 14	**138**, **106**
120˚	257, 176	242, 184	404, 84	15, 8	**147**, **92**
130˚	254, 174	227, 175	391, 64	27, 1	**137**, **110**
140˚	243, 176	240, 193	383, 86	3, 17	**140**, **90**
150˚	242, 200	229, 195	360, 82	13, 5	**118**, **118**
160˚	239, 210	240, 216	347, 65	1, 6	**108**, **145**
170˚	244, 208	243, 223	320, 85	1, 15	**76**, **123**
180˚	243, 217	243, 230	318, 90	0, 13	**75**, **127**

TABLE VIII
ESTIMATING TARGET POSITION, BEFORE AND AFTER CTR PREPROCESSING, FOR TANK DATA.

cylinder. After preprocessing, the target echo becomes weaker, resembling the signal strength of an aluminum target. The majority of the views, all but two orientations with the highest correlation with a PVC target, classify the object as aluminum cylinder; see Table IV. For the marina source images, the target is detected in only 12 orientations, 9 classified as aluminum and 3 as PVC. For the preprocessed data, the classification is consistent with the pool data.

3) Target Orientation: Before analyzing the results, it is noted that the ground truth data accuracy is restricted by the stability of the imaging system's platform. In the controlled environment for collecting the tank data, the platform could be rigidly secured. In contrast, the marina (field) is exposed to currents, changing tides, and boat traffic, all affecting camera platform positioning. We estimated errors as large as ± 30 [deg].

In a real survey potentially carried out by an ROV, this platform has to have the ability to remain in a constant position for a period of time, and to change the sonar's angle of attack based on its depth. In deeper waters, the ROV is a more stable platform that the cantilever beam secured on a floating barge near the sea surface, for our experiments.

Marina Data **Red: x-error > 60 pix, Blue: y-error > 60 pix, Green: x-error and y-error > 60 pix**					
Source Image Angle	*Actual (x,y) Position*	*Calculated (x,y) (Before)*	*Calculated (x,y) (After)*	*Error in (x,y) (Before)*	*Error in (x,y) (After)*
0˚	337, 281	344, 296	389, **385**	7, 15	52, **104**
10˚	321, 291	356, 301	339, **295**	35, 10	18, **94**
20˚	354, 284	341, **357**	329, **385**	13, **73**	25, **101**
30˚	367, 280	376, **363**	341, **385**	9, **83**	26, **105**
40˚	390, 278	367, **364**	351, **449**	23, **86**	39, **171**
50˚	390, 218	-	387, **439**	-	3, **221**
60˚	417, 220	-	**354**, **405**	-	**63**, **185**
70˚	442, 231	382, **402**	**356**, **452**	60, **171**	**86**, **221**
80˚	452, 240	**331**, **364**	**373**, **453**	**121**, **124**	**79**, **133**
90˚	470, 244	**356**, **354**	415, **450**	**114**, **110**	55, **206**
100˚	468, 234	**397**, **320**	**313**, **386**	**71**, **86**	**155**, **152**
110˚	465, 231	**377**, **403**	436, **317**	**88**, **172**	29, **86**
120˚	460, 211	**344**, **376**	**361**, **403**	**116**, **165**	**99**, **192**
130˚	440, 185	-	**328**, **75**	-	**112**, **110**
140˚	429, 180	-	**359**, **108**	-	**70**, **72**
150˚	417, 167	-	**329**, **46**	-	**88**, **121**
160˚	404, 159	400, **344**	378, 117	4, **185**	26, 42
170˚	385, 163	-	374, 151	-	11, 12
180˚	378, 168	-	365, 168	-	13, 0

TABLE IX
ESTIMATING TARGET POSITION, BEFORE AND AFTER CTR PREPROCESSING, FOR MARINA DATA.

Data Type	*Detected*	*Total Images*	*Detection Rate*
Tank (Before)	19	19	100.0%
Tank (After)	19	19	100.0%
Marina (Before)	12	19	63.2%
Marina (After)	19	19	100.0%

TABLE X
SUMMARY OF RESULTS IN ESTIMATING TARGET LOCATION, BEFORE AND AFTER CTR PREPROCESSING.

The complete results are presented in Tables V and VI, with a summary given in Table VII. We note that the maximum error for the tank and marina data is 140 [deg] and 110 [deg], respectively; but the true maximum error is modulo $\theta_T = 90$ [deg]. This holds true because the targets are symmetrical about their horizontal and vertical axes. The MARF-TM algorithm depends mainly on the matched template class to determine its orientation. Because the classification

Data Set:	*Tank (Before)*	*Tank (After)*	*Marina (Before)*	*Marina (After)*
Max Error in x	273	147	121	155
Max Error in y	54	145	185	221

TABLE XI
ERROR ANALYSIS IN LOCATING THE STEEL TARGET (IN UNITS OF [PIX]), BEFORE AND AFTER CTR PREPROCESSING.

and orientation estimation are coupled, an incorrect match between the source image material and template class can lead to a large error in orientation. For the tank source images after preprocessing, the target's return strength becomes weaker. This decrease in pixel intensity may cause the algorithm to establish a better match with the pool wall reflection. Not surprisingly, this leads to an erroneous estimate of the target orientation. From the summary in Table VII, it is noted that the algorithm determines the orientation of a detected object with a minimum success rate of 57.9% for both data sets.

4) Target Location: Once the MAFR-TM algorithm has determined the class and orientation of the detected object, the location within the source image can be computed. The algorithm searches through the maximum correlation array to identify the peak location C_{STmax} that corresponds to the specific class and orientation, determined in the previous steps. The results for the tank and marina data are presented in Tables VIII through XI.

The tank and marina images have a resolution of 575 [pix] $\times$ 568 [pix] and 800 [pix] $\times$ 568 [pix], respectively. An error within 60 [pix] means that the known location of the target is within approximately 90% in both x and y directions, providing a reasonable estimate of target position. In Tables VIII and IX, we have highlighted with red, blue and green font colors the location errors greater than 60 [pix] in the x, y and both directions, respectively.

For the tank source images after preprocessing, the large variations in the location of the true target is because the MAFR-TM algorithm matches the templates to the wall reflections, as is the case for the orientation estimation; see Fig. 19(a,a'). This is verified by the computed location. The true target is in the general area of 240-280 [pix] in the x-direction, and 155-220 [pix] in the y-direction, and the wall reflection covers a much larger area of 300-450 [pix] $\times$ 50-350 [pix], respectively. A summary of the overall performance in determining the proper target location has been summarized in Tables X and XI.

The algorithm performs well with the tank data before preprocessing, but the performance deteriorates significantly for the tank data after preprocessing and the marina data, both before and after preprocessing. Since the algorithm depends on the information from the previous steps, a large variance in the location can occur if the incorrect class and orientation are determined. That is, if the algorithm classifies the detected object incorrectly and/or specified an incorrect aspect angle, then the results for the actual location of the target would be skewed. Since the detection and classification performance of the algorithm decreased by 30% when applied to the marina data sets, the algorithm's localization performance is expected to decline as well.

TABLE XII
TARGET CHARACTERISTICS

Target	Description
θ_1	Cylindrical target with length l and radius $h/2$
θ_2	Cylindrical target with length l and radius h
θ_3	Cylindrical target with length $l/2$ and radius $h/2$
θ_4	Spherical target with radius $h/2$
θ_5	Spherical target with radius h

TABLE XIII
MASS ASSIGNMENT MODELS OF THE EVIDENCE GENERATED BY OPTICAL AND SONAR SENSORS. NOTE THAT, $A_1 = \{\theta_1, \theta_2, \theta_3, \theta_6\}$, $A_2 = \{\theta_4, \theta_5, \theta_6\}$, $A_3 = \{\theta_1, \theta_2, \theta_3\}$.

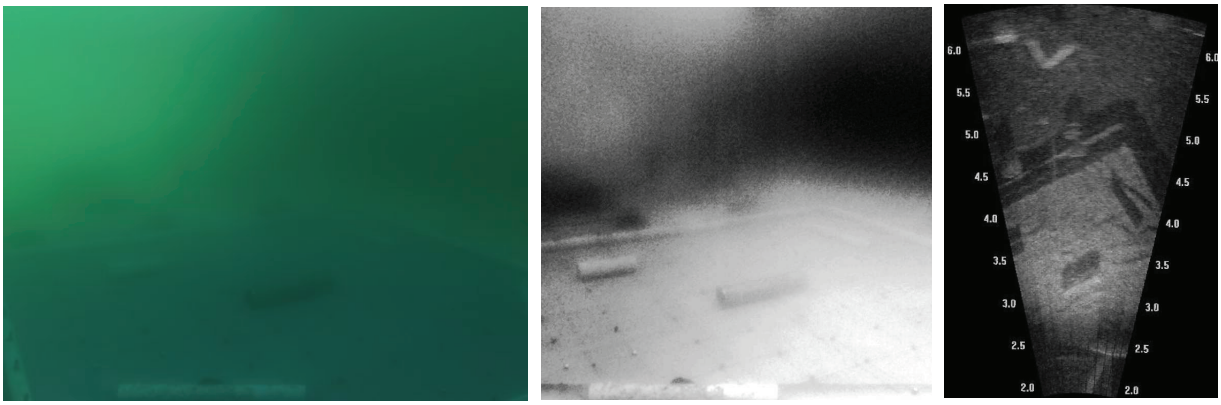
α	$\mathcal{E}_O(\alpha)$ $\{A_1, A_2, A_3, \Theta\}$	$\mathcal{E}_S(\alpha)$ $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \Theta\}$
0	$\{0.75, 0.0, 0.0, 0.25\}$	$\{.180, .139, .110, .039, .073, .459\}$
30	$\{0.75, 0.0, 0.0, 0.25\}$	$\{.189, .128, .099, .033, .067, .484\}$
60	$\{0.75, 0.0, 0.0, 0.25\}$	$\{.175, .132, .091, .042, .070, .490\}$
90	$\{0.0, 0.75, 0.0, 0.25\}$	$\{.099, .173, .098, .092, .165, .373\}$
120	$\{0.0, 0.0, 0.0, 1.0\}$	$\{.178, .126, .106, .041, .066, .483\}$
150	$\{0.0, 0.0, 0.75, 0.25\}$	$\{.185, .133, .114, .059, .090, .419\}$

For the marina data set, other elements can influence the results, including the target aging (biological growth, partial/complete burial, etc.), and the incorporation of clutter around the target. If the target appears partially buried or occluded from particular angle in some source images, these may not match the template of the true target, instead producing a better match between the template and a false target. Given acceptable results in many of the cases, fine tuning of the algorithm offers the potential for determining the detected target's position within 90% of the true location.

E. DS-based Fusion Technique

We employ the data acquired in the FAU outdoor pool, with the target O_Θ being imaged by the 1-Mpix digital optical camera and a DIDSON system operating at 1.8MHz, as these rotate around the target. The height of the equipment setup from the ground plane and the distance to O_Θ are recorded. we assume 5 possible target types, identified as θ_i, $i = \overline{1,5}$, with $\theta_6 \neq \theta_i$, $i = \overline{1,5}$, representing any other object.

Note how the target dimensions are chosen: 1) the cylindrical targets θ_1 and θ_2 are both l in length, and therefore cannot be differentiated in a perpendicular view. Similarly, θ_1 and θ_3 have the same radius, and so cannot be discriminated from views that are nearly orthogonal to their circular surfaces. In fact, if either object θ_1 or θ_3 is viewed with its circular side facing the camera, it cannot be differentiated from θ_3 and θ_4. This represents a scenario where different targets have the same 2-D visual appearance in one or more views depending on the orientation relative to the target. We explore how fusing (i.e., combining) evidence from different angles may lead to a more accurate classification of the target type.



(a) (b) (c)

Fig. 20. (a) Sample color image from a 1M-pixel optical camera. (b) Processed gray-scale image. While the processing removes a significant amount of clutter thus making it easier to identify the object under given conditions, optical images may not reveal much information under more turbid conditions. (c) Sonar image taken from DIDSON camera with $\alpha = 30$.

Let α denote the incident angle of sensors with respect to the target. Given α, target dimensions and distance to target from cameras and ground level, the acoustic shadow image $\mathcal{S}h_{\theta_i}(\alpha)$ corresponding to each target type θ_i is computed. To characterize the object O_Θ, both optical and acoustic images of the object at incident angles of $0, 30, 60, 90, 120, 150$ degrees are utilized; see Figs. 21 and 22.

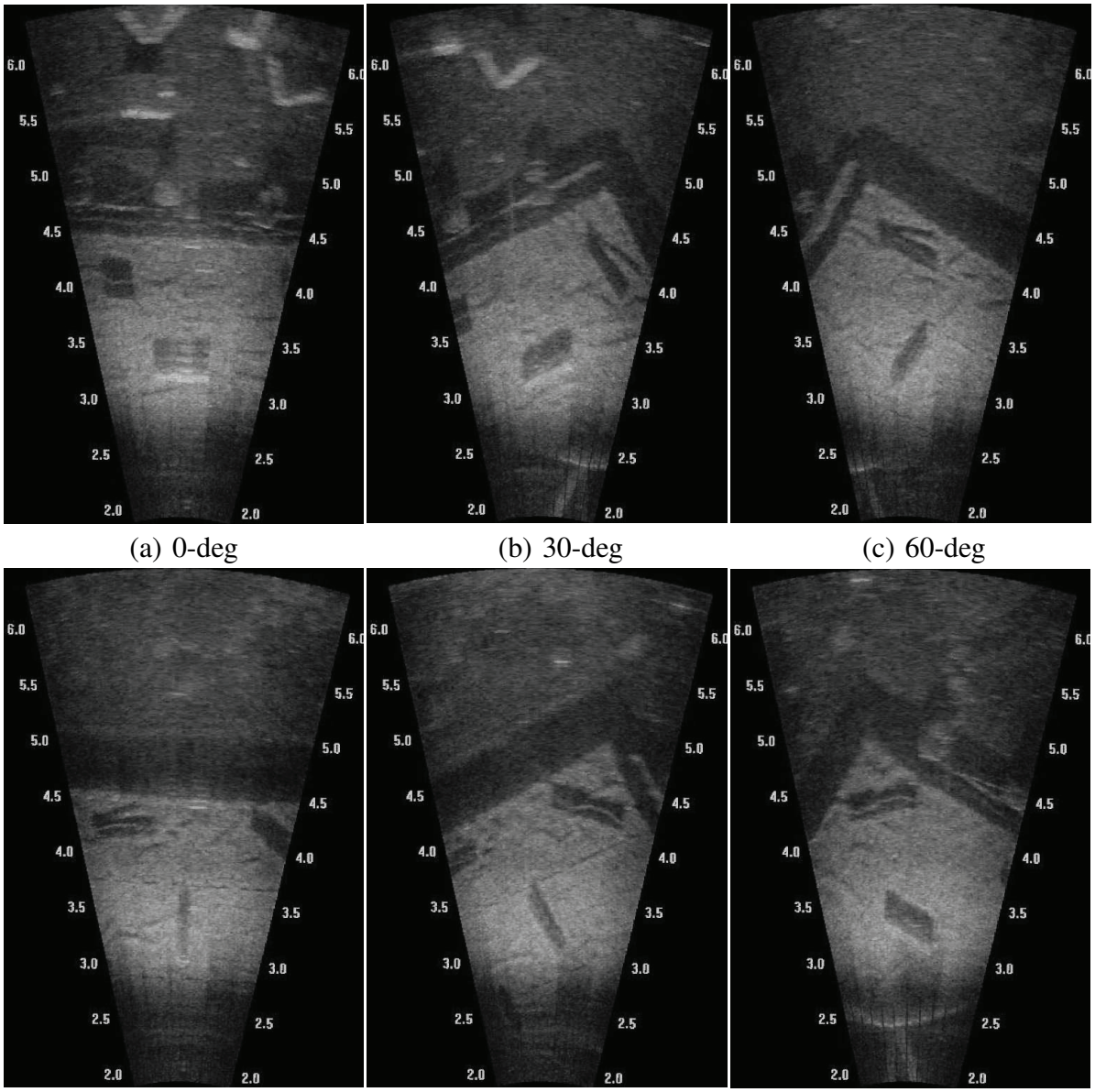
1) Image Preprocessing and Shadow Marking: Visual identification can become very difficult, if not impossible, in poor visibility; e.g., see the optical view of target in Fig. 20(a). Often, some image enhancement may be necessary to improve the image quality; see Fig. 20(b). In sonar images, the shadow boundaries serves as the visual cue for target classification; see section IV-B3; see Figs. 20(c) and 21. Here, the demarcation of the shadow from the object can sometimes become a challenging task due to factors that may affect both the intensity and geometry of the shadow and object regions. For the purposes of this demonstration, it suffices to identify the shadow region manually in acoustic images at different object orientations; see Fig. 21.

Figure 23(b) shows an acoustic image registered on the corresponding optical image. These registered images are useful in aiding the task of shadow selection.

2) Evidence Modeling: We generate $\mathcal{E}_O$ for each α by manually assigning the weights depending on the ability to visually identify targets. Note that, when it comes to simulating a human observer, only the object shapes can be identified. Inability to identify objects precisely in a given view is modeled by assigning a mass to the complete ambiguity Θ.

Remarks:

- We model a human observer with $\mathcal{E}_O$. Under low visibility conditions, we assume that the human observer can only identify the object shape, i.e., he/she can determine it is cylindrical or spherical with a certain level of confidence. If he/she cannot determine the shape, the



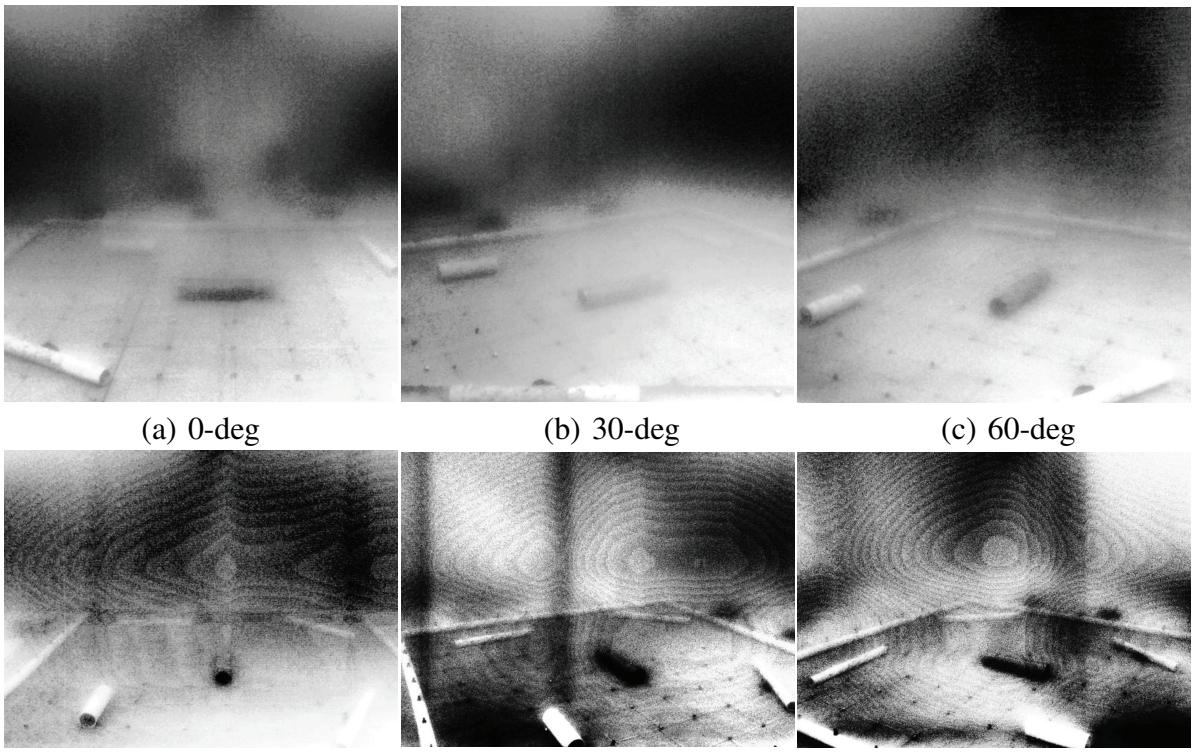
(a) 0-deg (b) 30-deg (c) 60-deg

(d) 90-deg (e) 120-deg (f) 150-deg

Fig. 21. Acoustic images of the object O_Θ at angles $0 - 150$ in steps of 30 degrees. Note the variation in shadow geometry with α.

object is identified as 'other' which is represented by θ_6.

- Take, for instance, the $\alpha = 0$ case. The object under observation appears to have a cylindrical object. So, this can be any of the target types θ_1, θ_2, θ_3 with specified dimensions, or any 'other' object (θ_6) that has a similar appearance to a cylinder in the current view. So, the observer assigns 75% confidence to $(\theta_1, \theta_2, \theta_3, \theta_6)$; the remainder is often assigned to Θ. This latter mass assigned to Θ can move into any $\theta \in \Theta$.



(a) 0-deg (b) 30-deg (c) 60-deg

(d) 90-deg (e) 120-deg (f) 150-deg

Fig. 22. Preprocessed optical images of the object O_Θ at angles $0-150$ at steps of 30 degrees. In views (b,c), it is clear that object is cylindrical. However, view (a) could well be some rectangular object. This uncertainty is even more pronounced in view (d).

Shadow analysis on acoustic images allows one to quantify evidence and generate confidence values to represent whether an observed object is of type $\theta_i \in \Theta$. This is done via a measure $\mu_{\theta_i}(\bullet) : \mathcal{S}h_{\theta_i}(\bullet) \times \mathcal{S}h_{O_\Theta}(\bullet) \mapsto [0,1]$, which captures the 'similarity' of $\mathcal{S}h_{\theta_i}(\bullet)$ to $\mathcal{S}h_{O_\Theta}(\bullet)$, i.e., the shadow image of the object under investigation. We assign the 'singleton' mass $m_S(\theta_i) = 0.2 \times \mu_{\theta_i}(\alpha)$, for each $\theta_i \in \Theta$; the remaining mass is allocated to Θ, for each α. See Table XIII for an example mass assignment for object O_Θ.

Remarks:

- In case of sonar views, we compute the similarity of the shadow of the imaged object (θ) to shadows of five target types.
- The shadow shape of θ ($\mathcal{S}h_{O_\Theta}$ represents the shadow of θ) is compared to shadows of the five known target types $\theta_1, \ldots, \theta_5$ ($\mathcal{S}h_{\theta_i}$, $i = \overline{1,5}$, represent the target shadows).
- In DS theory, mass assignment is highly subjective and application dependent. However, DS is very robust against mass assignment and modeling errors.

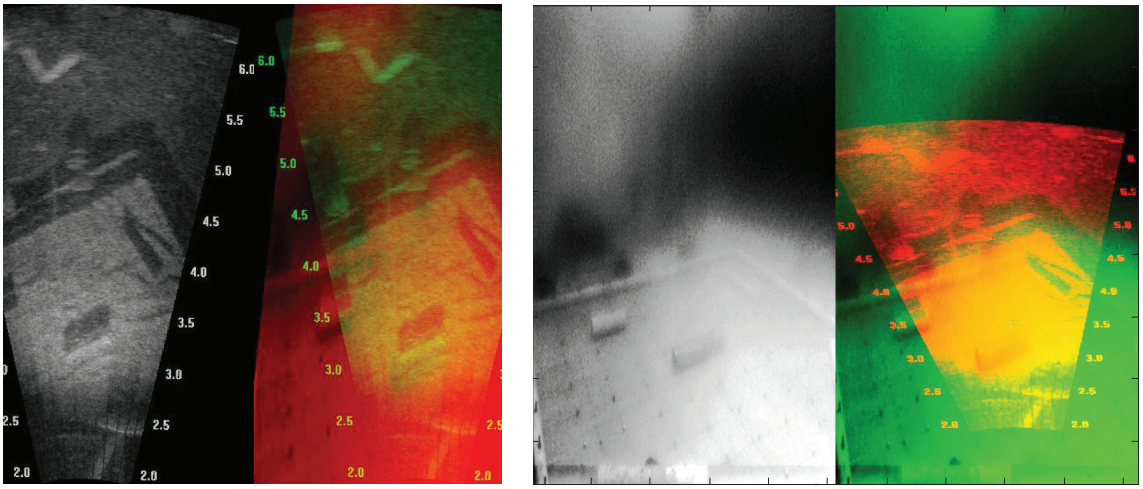


Fig. 23. (a) optical image on acoustic image, and (b) vice-versa. ($\alpha = 30$ image set.)

- The measure $\mu_{\theta_i}(\alpha)$ simply compares the shape of the shadow of target θ_i at angle α to the shape of the shadow of object under consideration. if the shadows are equivalent it will assign a value between 0 and 1 depending on whether shadow shape is being total disagreement or agreement.

3) Combining Evidence Sources: In the current setup, evidence fusion is not trivial. One can choose to combine/update evidence from individual sources over different α, or one can combine two different sources for a fixed α. We take a more intuitive approach to this with the help of the Dempster's Combination Rule (DCR) and Conditional Update Equation (CUE).

Optical Source Alone: Here we study the use of the optical source alone. One can rotate around a target to obtain as much information as possible to characterize the object. As one rotates around the object, one can update the existing belief about the object. This is illustrated under the column $\mathcal{E}_O(\theta - 1) \lhd \mathcal{E}_O(\theta)$ in Table XIV.

Observe the BoEs corresponding to $\alpha = 0, 30, 60$ (column #3). In all three cases, a significantly higher mass is assigned to $(\theta_1, \theta_2, \theta_3, \theta_6)$. This represents that the object seen is cylindrical $(\theta_1, \theta_2, \theta_3)$ or 'other'. So, initially, more evidence appears to support a target which is not spherical.

As more views are considered, the support moves more towards a cylindrical target as can be seen by the increased mass for $(\theta_1, \theta_2, \theta_3)$ at $\alpha = 150$; also see Fig. 22.

However, it appears that one cannot decide on one particular target type. Also, notice the fact that there is a significant amount of mass being assigned to Θ, representing the ambiguity in the available evidence. Thus, even after fusing 6 views together, it appears that one cannot make a crisp decision based on this evidence alone.

Acoustic Source Alone: Individual evidence corresponding to each α from the acoustic sensor

is very uncertain (Table XIII). For instance, $\mathcal{E}_S(60)$ (the BoE corresponding to the sonar evidence at $\alpha = 60$) leans more towards target type θ_1, whereas $\mathcal{E}_S(90)$ leans more towards target type θ_2. However, the uncertainty in both of these BoEs is quite high as indicated by the high mass for complete ambiguity.

Let us again explore the pattern of belief change as one updates the belief over different passes. One clear observation is that the mass assigned to the total ambiguity decreases over the iterations. This indicates that the uncertainty of the available evidence reduces as more views are combined.

The support towards target type θ_1 also appears to increase. However, the support towards θ_2 and θ_3 are not negligible given the mass assigned to total ambiguity.

Fused Optical and Acoustic Sources: The obvious fusion strategy is to combine evidence from each source for each α separately. The results obtained in this way are tabulated in column #1 of Table XIV. We have used DCR for combining $\mathcal{E}_S$ and $\mathcal{E}_O$. An alternate, more intuitive, fusion strategy is the following: at each observation, update the belief by first fusing evidence from both optical and acoustic sources together and then using this combined evidence to update the existing belief. See the column $(\mathcal{E}(\theta - 1) \lhd \mathcal{E}_O(\theta)) \lhd \mathcal{E}_S(\theta)$ in Table XIV for the BoEs obtained at each α.

Table XIV tabulates the fusion results for different fusion methods described earlier. Column #1 shows fused results obtained by combining evidence from each source for each α separately. This can be thought of as using only the information available at present. Columns #2 and #3 represent the updated masses of optical and sonar sources respectively. This is similar to updating one's belief using only one source of information. The last column shows a full update of information using both sonar and optical.

As one may notice, the observations are very similar to the previous cases in terms of the decreasing overall uncertainty. As more evidence is gathered, the support moves towards target type θ_1 and the overall uncertainty represented by the support for Θ decreases. Fig. 24 shows the change of belief of target type θ_1 as more views are fused together. The lower limit is the $Bl(\theta_1)$ and the upper limit is the $Pl(\theta_1)$; these correspond, respectively, to the lower and upper envelopes of any probability assignment compatible with the available evidence. The difference $Pl(\theta_1) - Bl(\theta_1)$ represents the associated uncertainty.

In summary, in the underwater environment which is rife with various types of data and evidence imperfections, the DS theoretic framework appears to be well suited to make decisions without making unrealistic and unjustifiable assumptions regarding these imperfections. In exploring the applicability of existing DS belief theoretic data fusion techniques to address the challenges involved in UXO characterization, we have illustrated the use of a recent DS theoretic evidence updating technique and the de facto DCR for the fusion of evidence obtained from optical and sonar image data. As the results demonstrate, in the absence of sufficient evidence, the DS theoretic framework does not make a firm commitment to any one object type, instead being receptive to further evidence that may enable for better discrimination. Moreover, the DS theoretic decisions incorporate confidence measures (in terms of belief and plausibility values)

TABLE XIV

FUSED EVIDENCE BOES. NOTE THAT, $A_1 = \{\theta_1, \theta_2, \theta_3, \theta_6\}$, $A_2 = \{\theta_4, \theta_5, \theta_6\}$, $A_3 = \{\theta_1, \theta_2, \theta_3\}$. HIGHEST AND SECOND HIGHEST MASSES IN EACH BOE, REPRESENTING THOSE PROPOSITIONS WITH THE HIGHEST 'SUPPORT', ARE MARKED IN **BOLD** AND <u>UNDERLINED</u>, RESPECTIVELY.

	Fused Evidence BoE $\mathcal{E}(\theta)$			
	$\mathcal{E}_O(\theta) \oplus \mathcal{E}_S(\theta)$	$\mathcal{E}(\theta-1) \lhd \mathcal{E}_O(\theta)$	$\mathcal{E}(\theta-1) \lhd \mathcal{E}_S(\theta)$	$(\mathcal{E}(\theta-1) \lhd \mathcal{E}_O(\theta)) \lhd \mathcal{E}_S(\theta)$
$\alpha \downarrow$	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, A_1, A_2, A_3, \Theta\}$	$\{A_1, A_2, A_3, \Theta\}$	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \Theta\}$	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, A_1, A_2, A_3, \Theta\}$
0	$\{.196, .152, .120, .011, .020, \underline{.376}, .000, .000, \mathbf{.501}\}$	$\{\mathbf{.750}, .000, .000, \underline{.250}\}$	$\{\underline{.179}, .139, .110, .039, .073, \mathbf{.460}\}$	$\{.131, .102, .080, .029, .054, \mathbf{.375}, .000, .000, \underline{.229}\}$
30	$\{\underline{.204}, .139, .107, .009, .018 \mathbf{.392}, .000, .000, .131\}$	$\{\mathbf{.844}, .000, .000, \underline{.156}\}$	$\{\underline{.230}, .165, .128, .044, .086, \mathbf{.347}\}$	$\{\underline{.181}, .129, .100, .029, .058, \mathbf{.358}, .000, .000, .145\}$
60	$\{\underline{.191}, .144, .099, .011, .019 \mathbf{.402}, .000, .000, .134\}$	$\{\mathbf{.891}, .000, .000, \underline{.109}\}$	$\{\underline{.246}, .181, .132, .053, .095, \mathbf{.293}\}$	$\{\underline{.198}, .146, .106, .034, .061 \mathbf{.352}, .000, .000, .103\}$
90	$\{.034, .060, .034, .128, \underline{.228} .000, \mathbf{.387}, .000, .129\}$	$\{\underline{.445}, \mathbf{.469}, .000, .086\}$	$\{.190, \underline{.209}, .133, .090, .161, \mathbf{.217}\}$	$\{.140, .144, .093, .071, .127 \underline{.176}, \mathbf{.180}, .000, .069\}$
120	$\{\underline{.178}, .126, .105, .041, .066 .000, .000, .000, \mathbf{.484}\}$	$\{.223, \underline{.234}, .000, \mathbf{.543}\}$	$\{\mathbf{.227}, .198, .145, .075, .130, \underline{.225}\}$	$\{\underline{.177}, .148, .110, .060, .103 .088, .090, .000, \mathbf{.224}\}$
150	$\{\underline{.209}, .150, .128, .017, .250 .000, .000, \mathbf{.354}, .117\}$	$\{.111, .117, \mathbf{.469}, \underline{.303}\}$	$\{\mathbf{.245}, .193, .153, .080, .129, \underline{.200}\}$	$\{\mathbf{.196}, .151, .121, .056, .090 .044, .045, \underline{.163}, .134\}$

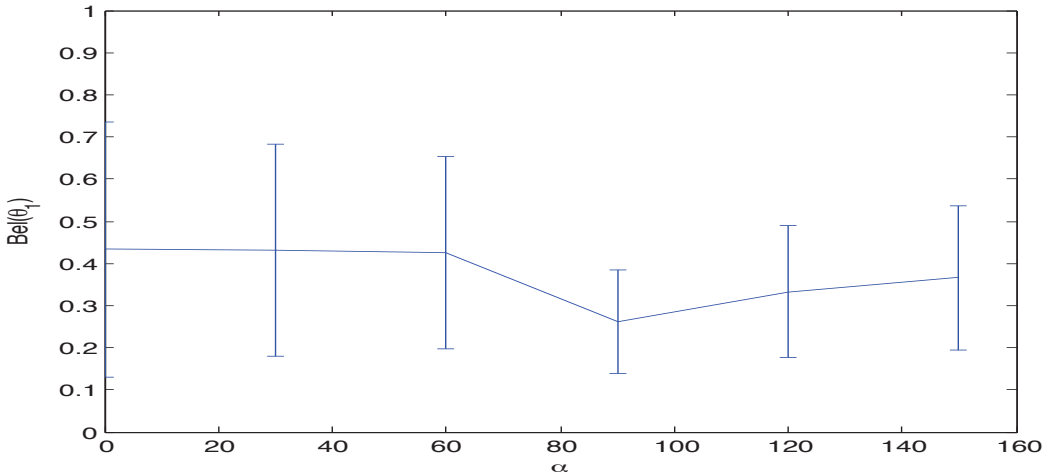


Fig. 24. Change in belief of target type θ_1 as more and more views are combined together.

that provide invaluable information for a domain expert to make an informed decision.

The evidence modeling and shadow analysis could be further improved to enhance the performance. Incorporation of additional sensor modalities and complementary classification techniques into the fusion framework is expected to improve the overall performance. In particular, we conjecture that a mosaic—a global view of the site (say, a debris field) that is constructed from

a large set of registered views—can offer a richer visual cue about the scene content than single isolated views. As we have described, optical and multi-modal overlapping images, when treated as stereo views, enable the 3-D reconstruction of objects. In particular, with as little information as three correspondences in optical and sonar views, we can construct the contours of various targets in 3-D from their shadows in the sonar image.

This one-year project did not afford us sufficient time to explore a detailed investigation of how to incorporate global site information from mosaics, as well as to develop and test a theory-based formulation for exploiting the 3-D structure and (or) occluding contours.

VII. Recommendations and Future Work

A. *Optical Imaging*

Undoubtedly, optical imaging is the most mature and natural imaging modality for target classification and identification. We believe that optical imaging must be an integral component of UXO identification systems. However, the serious bottleneck is the visibility limitation due to high turbidity within "harsh environments." To be deployed as a reliable sensing and mapping modality, improved imaging strategies have to be explored, tested and implemented. Among potential technologies are intelligent active lighting and sensing including scanning laser, as well as scene radiance restoration methods. Any of these solutions attempt to overcome the backscatter, which is the most significant source of image degradation.

This has motivated recent activities to devise new stereo techniques for 3-D target reconstruction within scattering media, including underwater. One approach we have undertaken in achieving robust 3-D reconstruction is to make use of both the polarization and stereo depth cues [28]. While quiet effective, this particular method relies on the calibration of the sources in active artificial lighting, as well as the medium optical properties. Very recently, we have devised another technique that requires little calibration. Here, all needed is the backscatter field at infinity; namely, an image with no target in the field of view [29]. We have obtained promising results in controlled water tank experiments, with ocean tests underway. Our method's novelty, unlike earlier methods for underwater stereo that aim to remove the backscatter, is that we actually exploit the depth cue in the backscatter field (structured noise component), in addition to the traditional binocular disparity as the signal component. An advantage of our method is that we simultaneously achieve dense scene reconstruction and enhanced de-scattered views; that is, images where the scatter field has been removed. Fig. 25(a) depicts a sample stereo pair acquired in our water tank. In (b), we have given the left view, overlaying a diagonal section with the enhanced section to facilitate the comparison over all three objects.

As stated, the laser line scan (LLS) serial imaging technique is another effective technology to combat backscattering. This is accomplished by reducing the common volume between the sensor's instantaneous viewing angle cone and the illumination volume of the artificial source which are scanned in a synchronous fashion, intersecting at the target plane. In this way, the LLS system utilizes narrow spatial and angular filtering to reject undesirable scattering, while providing wide total field of view imagery. Here, a narrow blue-green laser beam is swept in

(a) (b)

Fig. 25. (a) Stereo views in scattering media with magnified view of section over a cylindrical target before and after enhancement. (b) Overlaying a diagonal section covering all three objects to facilitate the comparison.

one direction across the scene over the region that is being imaged (e.g., ocean bottom) and the reflected light generates a gray-scale image. The LLS system is towed across the target in the perpendicular direction (to the scanning direction) at some fixed distance by some submersible platform (e.g., an AUV or ROV) to produce the 2-D image. Generally, LLS systems have an operational range of up to five times the range of conventional photography cameras (CCDs and recent digital cameras). The main drawbacks is their bulky size and narrow operational envelope (performance is sensitive to small variations in turbidity, narrow depth of field), but making the packaging more and more compact while improving the robustness of existing bulky systems will increase their deployment potential significantly.

B. FS Sonar Imaging

We have concluded from this study that high-frequency 2-D FS sonar systems can provide and effective technology for UXO detection, and significantly enhance the classification capability. Having said this, fundamental research is necessary to develop robust algorithms to experiment with a larger database of targets and environmental conditions. For example, it is well-known and our study has verified that acoustic signature of sought after targets can vary with pose relative to the sonar, and particularly in the presence of clutter, e.g., other objects with similar reflectance properties. Given the diverse characteristics of existing FS sonar systems, the development of an effective detection and classification system requires a theoretical physics-based approach that accounts for target characterization, environmental conditions, as well as sonar properties.

It is important to note that the FS sonar imaging technology is only steps past the infancy stage. In particular, while originally developed as a tool primarily for human visualization in highly turbid environments with minimal/no visibility, hardware and design improvements to reduce noise and distortion will facilitate the application of scene reconstruction methods to make more accurate position and target measurements.

C. DS Theoretic Evidence Fusion

The underwater environment is rife with various types of data and evidence imperfections that makes it enormously difficult, or sometimes even impossible, to apply more 'traditional' techniques for UXO identification and classification purposes. The preliminary work conducted during this one-year seed project appears to indicate that the ambiguity and imperfection handling mechanisms intrinsic to the DS theoretic framework are better suited in such an environment. This is mainly because the DS theoretic framework requires little or no recourse to the types of unrealistic and unjustifiable assumptions that encumber the traditional approaches.

As the results demonstrate, our DS theoretic strategy refrains from making a firm commitment towards any one object type when sufficient evidence is lacking. Instead, by having non-zero support allocated to the completely ambiguous proposition Θ, the strategy remains receptive to further evidence that may enable for better discrimination. Perhaps more importantly, the DS theoretic decisions incorporate confidence measures (in terms of belief and plausibility values) that provide invaluable information for a domain expert to make an informed decision while having full knowledge of the underlying uncertainties.

The scope of our work was restricted to evidence obtained from optical and sonar image data and the fusion of this data via the widely used DCR and a recent DS theoretic evidence updating technique. In light of the study we have undertaken and the lessons learned from it, we make several promising avenues of research and exploration.

1) Shadow Analysis: Our preliminary study shows great potential for using object shadows as "signatures" for the purpose of identification and classification of underwater UXOs. To obtain improved classification performance, sophisticated and more effective shadow models need to be investigated.

2) Evidence Filtering: Real-time processing and fusion of (potentially imperfect) streaming data generated from heterogeneous sources can be accomplished via evidence filters (EFs) that operate directly on DS theoretic evidence while offering the power, efficiency, and convenience of digital signal processing. EFs have known to be quite effective for fusing multi-modal evidence to detect faint signatures that are potentially buried in clutter [9]. This appears to be exactly what is needed in underwater UXO detection and classification.

3) Additional Sensor Modalities and Complementary Classification Techniques: While our one-year seed study allowed us exploration of only two modalities, incorporation of other modalities (e.g., object contours) into the fusion process is crucial for better performance. For example, how does one incorporate the richer visual cues one gets from a mosaic or a 3-D reconstruction of objects generated from their shadows? Recent work that has led to significant computational improvements in the conditional approach to DS theoretic evidence updating and fusion is expected to make this task more feasible. This an issue that requires further study.

4) Reliability Issues: Reliability of evidence sources must be taken into account when multiple sources are fused. For example, optical sources provide more accurate measurements in shallow

depth levels, but its usefulness decreases at increased depths. On the other hand, sonar imaging becomes less reliable under certain environmental conditions (e.g., in areas having increased levels of reflections). Capturing this reliability information calls for development of techniques for modeling domain expert opinions and then incorporating these models into the fusion process itself.

D. Other Issues

While this study has concentrated on the sensing issues, the close coupling with platform navigation should not be ignored. A stable platform (ROV or UUV) with capability to execute precise navigation commands for maintaining desired poses relative to potential targets plays a critical role in acquiring data with the highest quality and information content. Now, desired platform poses, both position and orientation, are not known ahead of time in most cases. Unless assessed entirely by a human operator in the loop, partial or fully automated operations require in situ processing, and therefore the implementation of data processing in real-time embedded processors. Additionally, the processing of visual data enables target-based positioning, which for our application, is a more critical capability than absolute positioning, including for carrying out target reacquisition.

VIII. Publications

- Professional Meetings
  1) S. Nagahdaripour, P.P. Beaujean, L. Brisson, K. Premartne, T. L. Wickramarathne, "Belief Theoretic Multi-Sensory Data Fusion for Underwater UXO Identification," Poster at Partners in Environmental Technology Technical Symposium & Workshop, Washington D.C. (December 2009).
  2) L. N. Brisson, P.P. Beaujean, S. Nagahdaripour, "Multiple-Aspect Fixed-Range Template Matching for the Detection and Classification of Underwater Unexploded Ordnance in DIDSON Sonar Images," *Proc. MTS/IEEE Oceans'2010 conference*, Seattle, Washington (September 2010).
  3) S. Negahdaripour, H. Assalih, Y. Petillot, and L.N. Brisson, "Performance and accuracy in visual motion computation from FS sonar video sequences," *Proc. MTS/IEEE Oceans'2010 conference*, Seattle, Washington (September 2010).
  4) T. Wickramarathne, K. Premaratne, S. Negahdaripour, Lisa Brisson, and P.P. Beaujean, "A Belief theoretic approach for characterization of underwater munitions," *Proc. MTS/IEEE Oceans'2010 conference*, Seattle, Washington (September 2010).
- Dissertation
  1) S.M. Hempel, "Multimodal image registration; application to acoustic and optical imaging in scattering media," M.S. thesis, Technical University of Munich, Munich, Germany (November 2009).
  2) L.N. Brisson, "A Methodology to Detect and Classify Underwater Unexploded Ordnance in DIDSON Sonar Images", M.S. thesis, Florida Atlantic University, Florida (May 2010).

REFERENCES

[1] AETC Inc. "Marine Towed Array Technology Demonstration at the Former Naval Duck Target Facility," ESTCP Project #200324, 2005.

[2] H. Altincay, "A Dempster-Shafer theoretic framework for boosting based ensemble design," *Pattern Analysis and Applications Journal*, vol. 8(3), pp. 287–302, Dec. 2005.

[3] H. Altincay and M. Demirekler, "Speaker identification by combining multiple classifiers using Dempster-Shafer theory of evidence," *Speech Communication*, vol. 41 (4), pp. 531–547, Nov. 2003.

[4] S. S. Anand, D. A. Bell, and J. G. Hughes, "EDM: A general framework for data mining based on evidence theory," *Data and Knowledge Engineering*, vol. 18, pp. 189–223, 1996.

[5] P. L. Bogler, "Shafer-Demspter reasoning with applications to multisensor target identification systems," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 17(6), pp. 968–977, 1987.

[6] J. R. Boston, "A signal detection system based on Demspter-Shafer theory and comparison to fuzzy detection," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 30 (1), pp. 45–51, Feb. 2000.

[7] D. Cai, M. F. McTear, and S. I. McClean, "Knowledge discovery in distributed databases using evidence theory," *International Journal of Intelligent Systems*, vol. 15 (8), pp. 745–761, June 2000.

[8] T. M. Chen and V. Venkataramanan, "Demspter-Shafer theory for intrusion detection in ad hoc networks," *IEEE Internet Computing*, vol. 9(6), pp. 35–41, Nov./Dec. 2005.

[9] D. A. Dewasurendra, P. H. Bauer, and K. Premaratne, "Evidence filtering," *IEEE Transactions on Signal Processing*, vol. 55(12), pp. 5796–5805, 2007.

[10] F. Delmotte and P. Smets, "Target identification based on the transferable belief model interpretation of Dempster-Shafer model," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 34(4), pp. 457–471, July 2004.

[11] Z. Elouedi, K. Mellouli, and P. Smets, "Belief decision trees: theoretical foundations," *International Journal of Approximate Reasoning*, vol. 28(2/3), pp. 91–124, Nov. 2001.

[12] R. Fagin and J. Y. Halpern, "A new approach to updating beliefs," in *Proc. Conference on Uncertainty in Artificial Intelligence (UAI'91)*, P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, Eds. New York, NY: Elsevier Science, pp. 347–374, 1991.

[13] S. L. Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35(4), pp. 1018–1031, July 1997.

[14] G. K. Hewawasam, K. Premaratne, and M.-L. Shyu, "Rule mining and classification in a situation assessment application: a belief theoretic approach for handling data imperfections," *IEEE Transaction on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 37(6), pp. 1446–1459, Dec. 2007.

[15] F. Janez and A. Appriou, "Theory of evidence and non-exhaustive frames of discernment: Plausibilities correction methods," *Int. Journal of Approximate Reasoning*, vol. 18(1/2), pp. 1–19, Jan./Feb. 1998.

[16] E. C. Kulasekere, K. Premaratne, D. A. Dewasurendra, M.-L. Shyu, and P. H. Bauer, "Conditioning and updating evidence," *International Journal of Approximate Reasoning*, vol. 36(1), pp. 75–108, Apr. 2004.

[17] H. Leung and J. Wu, "Bayesian and Dempster-Shafer target identification for radar surveillance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36(2), pp. 432–447, Apr. 2000.

[18] M. Lhuillier, and L. Quan, "Match propagation for image-based modeling and rendering," *IEEE T. Pattern Analysis Machine Intelligence*, Vol 24 (8), pp. 1140–1146, 2002.

[19] J. McDonald J, "Technology Needs for Underwater UXO Search and Discrimination," Final Report of SERDP Project UX-1322, 2004.

[20] S. Negahdaripour, "Revised interpretation of optical flow; Integration of radiometric and geometric cues in optical flow," *IEEE Trans. PAMI*, Vol 20(9), pp. 961-979, September, 1998.

[21] S. Negahdaripour, H. Sekkati, H. Pirsiavash, "Opti-acoustic stereo imaging: On system calibration and 3-D target reconstruction," *IEEE Trans. Image Processing*, *IEEE Trans. Image Processing*, Vol 18(6), pp. 1203-1214, June, 2009.

[22] S. Negahdaripour, A. Taatian, "Integration of motion cues in 2-D optical and sonar video for 3-D positioning," *Computer Vision Image Understanding* Vol 114(8), pp. 928-941, August, 2010.

[23] G.P. Penney, J. Weese, J.A. Little, P. Desmedt, D.L.G. Hill, and D.J. Hawkes, "A comparison of similarity measures for use in 2-D - 3-D medical image registration, *IEEE Transaction on Medical Imaging* Vol 16(2):187198, 1997.

[24] S.L. Rose-Pehrsson, K. Johnson, C. Minor, and V. Guthrie, "Intelligent Data Fusion for Wide-Area Assessment of UXO Contamination," SERDP Project MM-1510 2006 Annual Report NRL/MR/6180-07-9039, Naval Research Laboratory, Washington, DC, 2007.

[25] K. Premaratne, D. A. Dewasurendra, and P. H. Bauer, "Evidence combination in an environment with heterogeneous sources," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 37(3), pp. 298–309, 2007.

[26] K. Premaratne, M. N. Murthi, J. Zhang, M. Scheutz, and P. H. Bauer, "A Dempster-Shafer theoretic conditional approach to evidence updating for fusion of hard and soft data," in *Proc. International Conference on Information Fusion (ICIF'09)*, Seattle, WA, pp. 2122–2129, July, 2009.

[27] R.P. Reid, D. Lirman, N. Gracias, S. Negahdaripour, A. Gleason, and B. Gintert, "Application of landscape mosaic technology to complement coral reef resource mapping and monitoring," Final report of SERDP project SI-1333, March 2010.

[28] A. Sarafraz, S. Negahdaripour, Y. Schechner, "Enhancing images in scattering media utilizing stereovision and polarization," *Proc. WACV'09*, Snowbird, UT, December, 2009.

[29] A. Sarafraz, S. Negahdaripour, Y. Schechner, "Performance assessment in solving the correspondence problem in underwater stereo imagery," *IEEE/MTS Oceans'10 Conference*, September, 2010.

[30] *SERDP and ESTCP Workshop on Technology Needs for the Characterization, Management, and Remediation of Military Munitions in Underwater Environments*, Final Report, October, 2007.

[31] M. Sezgin and B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation," in *Journal of Electronic Imaging* Vol 13(1), pp. 146165, 2003.

[32] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press, 1976.

[33] T. L. Wickramarathne, K. Premaratne, M. N. Murthi, and M. Scheutz, "A Dempster-Shafer theoretic evidence updating strategy for non-identical frames of discernment," in *Proc. Workshop on the Theory of Belief Functions (WTBF'10)*, Brest, France, Apr. 2010.

[34] T. L. Wickramarathne, K. Premaratne, M. Kubat, and D. T. Jayaweera, "CoFiDS: A belief theoretic approach for automated collaborative filtering," *IEEE Transactions on Knowledge and Data Engineering*, May 2010.

[35] T.L. Wickramarathne, Kamal Premaratne, "Focal elements generated by the dempster-shafer theoretic conditionals: A complete characterization," in *Proc. International Conference on Information Fusion (ICIF'10)*, Scotland, UK, July 2010.

[36] *International Conference on Information Fusion (ICIF'09)*. Seattle, WA: International Society of Information Fusion, July 2009.

[37] R. R. Yager, M. Fedrizzi, and J. Kacprzyk, Eds., *Advances in the Dempster-Shafer Theory of Evidence*. New York, NY: John Wiley and Sons, 1994.

Appendix A: Details of MAFR-TM Algorithm

A.1 - Overview

The Multiple-Aspect Fixed-Range Template Matching (MAFR-TM) algorithm is a sonar-specific methodology to detect and classify UXO using the low resolution sonar data captured by the DIDSON US300 (it can be easily extended to the BlueView sonar data as well). This technique is designed to detect and classify a target of high characteristic impedance in an environment that contains similar shaped objects of low characteristic impedance. The MAFR-TM algorithm is based on the proven concept of the template matching technique, which is a two-dimensional correlation between a reference image (template) and an image collected during field operations (source image). In the MAFR-TM algorithm, the template matching method is efficiently implemented in the wave number domain using two-dimensional Fast Fourier Transforms (2D-FFT) and wave number leakage is reduced with an optimized separable two-dimensional Kaiser window.

MAFR-TM relies on the fact that UXO detection and classification is a re-acquisition process: the sonar is expected to scan a suspicious area, where UXO is likely to be present. Hovering vehicles (remotely piloted or autonomous) equipped with sector scan-sonar can follow a circular trajectory and create a 360 [deg] view of a small area (on the order of 20 [sqm]). MAFR-TM relies on a known vehicle trajectory to generate the templates under ideal conditions: in a tank with no clutter, limited bottom roughness and in the presence of a single target, at a known aspect angle and range. The templates are grouped into classes, each class representing a specific object (in this case, cylinders) observed at various aspect angles. Therefore, MAFR-TM does not need the two-tiered training-and-classification process of traditional computer-aided-detection-and-classification software. Instead, it can use a known database and perform the detection and classification "on the fly", so long as the vehicle keeps the potential target within a certain distance window. The platform does not have to follow exactly the same trajectory for MAFR-TM to work, as this technique defines areas where the target is likely to be present. The MAFR-TM algorithm is implemented in MATLAB and can be summarized as follows:

- Load a source image into workspace for orientations $k_S = \overline{1, K_S}$.
- Divide the source image by its standard deviation σ_S.
  - Initialize the template material, m_T.
    - ∗ Initialize the template orientation k_T.
    - ∗ – Load a template into the workspace for $k_T = \overline{1, K_T}$;
    - ∗ – Divide the template by its standard deviation σ_T;
    - ∗ – Calculate correlation between source image and template C_{ST};
    - ∗ – Normalize the correlation with respect to maximum value C_{STNorm};
    - ∗ – Define a probable target zone A_H by applying a high intensity threshold T_H to the normalized correlation;

    - * – Compute the vertical gradient of the source image ∇_V;
    - * – Acquire a probable shadow zone A_S by applying a threshold T_G to the vertical gradient along the range axis;
    - * – Calculate the minimum distance between the probable target and shadow zones d_{min}.
    - * – Apply the distance threshold T_d to the minimum distance computed above to determine if an object has been detected within the source image.
    - * – Determine the source image location (i_T, j_T) of the peak C_{STmax}, and store in memory.
    - * – Compute the signal-to-noise ratio (SNR) of the correlation and store in memory.
    - * Update the template orientation k_T.
  - – Update the template material m_T.
- Update the source image orientation k_S.
- Make the solution decision (class, orientation, and location) according to the results computed within the innermost loop.

These fundamental steps are repeated for the various template orientations and classes with every source image of a 360 [deg] survey. Once every source image of a 360 [deg] survey has been processed, the object is classified if the peak SNR corresponds to the same class across all (or a high percentage of) the aspect angles in every source image of the survey. The detected object's orientation is then identified by choosing the template that produced the maximum SNR for the current source image. Finally, the object's true position is determined based on the prior results of the algorithm. These essential concepts are presented in detail as follows.

A.2 - Source Image and Templates

Each source image S is an intensity plot of a target imaged at a fixed range of 3 [m] and at one of 36 different orientations (0 [deg] to 350 [deg] at 10 [deg] increments). Each pixel in the source image has coordinate (x, y) with an intensity range [0-255], where $x = \overline{1, X}$ and $y = \overline{1, Y}$; X and Y are the horizontal and vertical pixel dimensions of the source image, respectively. The x coordinate increases from left to right while y increases from top to bottom. This left hand coordinate system is used because images are stored in MATLAB as matrices, where the x-coordinate corresponds to the column index and the y-coordinate represents the row index of the source image. The source image is normalized with respect to its standard deviation,

$$\sigma_S = \sum_{x=1}^{X} \sum_{y=1}^{Y} \left(\frac{(s(x,y) - \bar{s}(x,y))^2}{XY - 1} \right)^{1/2}$$

where $s(x, y)$ in the source image, and $\bar{s}(x, y)$ is the mean pixel value within 3×3 region centered at location (x, y).

Three types of targets (steel, aluminum, and PVC) were used to create the templates. These targets were imaged at the same fixed range of 3 [m] and 36 different aspect angles ([0-360

[deg]] at 10 [deg] increments). One set of templates for each target class corresponds to 36 "ideal" sonar images. Each template T is also an intensity plot of the various targets with pixel value t. The template has the (m, n) coordinate system with intensity range [0-255], $m = \overline{1, M}$ is the column index, $n = \overline{1, N}$ is the row index, and M and N are the horizontal and vertical pixel dimensions of a template, respectively. The templates are also normalized with respect to their standard deviation,

$$\sigma_T = \sum_{m=1}^{M} \sum_{n=1}^{N} \left(\frac{(t(m,n) - \bar{t}(m,n))^2}{MN - 1} \right)^{1/2}$$

where $t(m, n)$ and $\bar{t}(m, n)$ are the template, and its mean value within 3×3 region centered at location (m, n), respectively.

The template T is then placed on top of the source image S so that the top left corner of the template is at the top-left corner of the source image. The template moves incrementally, pixel-by-pixel, from the left to the right and from the top down following the left hand coordinate system. In order to relate the source image and template coordinate systems, a new coordinate system (x', y') is introduced, where $x' = i\Delta m$ and $y' = j\Delta n$, and (i, j) denoted the position of the template relative to the top-left corner of the image.

A.3 - Correlation

Throughout the technique, the correlation is computed between each source image over every template, and therefore can be represented as a function of the position within the source image (x, y), the orientation of the source image θ_S, the orientation of the template θ_T, and the template material type m_T, and is computed for each new location (i, j):

$$C_{ST}(i, j) = C(i, j, k_S \Delta\theta_s, k_T \Delta\theta_T, m_T)$$

where $i = \overline{1, I}$ and $j = \overline{1, J}$ are the column and row indices for the correlation measure, $k_S = \overline{1, K_S}$ and $k_T = \overline{1, K_T}$ are the source and template image orientation indices, $\Delta\theta_s$ and $\Delta\theta_T$ are the angle increment for the source and template images (here, both are 10 [deg]), and $m_T = \overline{1, M_T}$ is the template material type index.

The correlation between one template and the portion of the source image the template overlaps is,

$$C_{ST}(i, j) = \sum_{m=1}^{M} \sum_{n=1}^{N} t(m, n) s(m + i, n + j)$$

If a peak C_{STmax} occurs in the correlation function, then an object similar to that of the template is detected in the source image S. The maximum value is defined by

$$C_{STmax}(i_{max}, j_{max}) = Max\{C_{ST}(i, j) | i = \overline{1, I}, j = \overline{1, J}\}$$

corresponding to the best match between the template and the source image, where (i_{max}, j_{max}) is the location of C_{STmax} in the source image. Where there is an ambiguity in establishing if a

true target has been detected, this is resolved based on the minimum acceptable distance between the target and shadow zones.

A.4 - Detection and Initial Localization

To reveal a probable location of an object within the field of view, one must examine the correlation for a single template scan. This probable location, or target zone, can be found by taking the correlation and normalizing it with respect to the highest peak:

$$C_{STNorm}(i,j) = \frac{C_{ST}(i,j)}{C_{STmax}}$$

where C_{STmax} is the maximum correlation over all the templates at location (i_{max}, j_{max}). Furthermore, the normalized correlation is thresholded by T_H to acquire the collection of points that describe the target zone, or

$$A_H = \{(i,j)|C_{STnorm}(i,j) \geq T_H\} \quad (i_H, j_H) \in A_H$$

The normalized correlation $C_{STNorm}(i,j)$ is also used in finding the shadow zone. The shadow zone is determined by calculating the vertical gradient of the correlation:

$$\nabla_V(i,j) = |C_{STNorm}(i,j+1) - C_{STNorm}(i,j)|$$

Another threshold T_G is applied to this region to acquire an approximate location of the target's shadow, or the area that describes the shadow zone:

$$A_S = \{(i,j)|\nabla_V(i,j) \geq T_G\} \quad (i_S, j_S) \in A_s$$

The distance between these two zones is then computed by

$$d = \sqrt{(i_h - i_s)^2 + (j_h - j_s)^2}$$

Due to the geometry and nature of sonar image image acquisition, the shadow is always located behind the target in every source image. Therefore, only the positive vertical gradient is considered to eliminate any spurious shadow zones. Furthermore, the elements of the shadow zone must have coordinates above the target zone's coordinates in the correlation. Using these conditions, the minimum distance between the target and shadow zones is defined as

$$d_{min} = Minimum\{d|\nabla_V(i_s, j_s) > 0,\ j_S < j_H\}$$

The decision whether an object has been detected in the source image is based on a distance threshold T_d, which is chosen according to the smallest template size in the entire template database; that is

$$T_d = Minimum\{X_{ROI,T}, Y_{ROI,T}\}$$

where $(X_{ROI,T}, Y_{ROI,T})$ is the length and width of a template. An object is detected if the minimum distance is less than the threshold distance T_d. The location of the detected object for a particular template scan can is determined from

$$(i_T, j_T) = \{(i, j)|d_{min} < T_d\}$$

where (i_T, j_T) is the correlation index along the length and width of the source image.

A.5 - Classification and Orientation

The SNR is utilized to identify the proper class of the target ($m_T = 1, 2, 3$ for steel, aluminum, and PVC, respectively) by comparing the maximum values of the correlation between template classes over every source image in a 360 [deg] survey. The SNR values are sorted according to classes and aspect angles. A target is considered valid and classified if the peak SNR corresponds to the same class across all (or a high percentage of) the aspect angles of an entire 360 [deg] survey. The SNR is defined here as the ratio of the peak area of the correlation to the average correlation, excluding the peak value, or

$$SNR = 10 * \log_{10}\left(\frac{C_{ST,H}}{C_{ST,N}}\right)$$

where $C_{ST,H}$ is the peak of the area of the correlation:

$$C_{ST,H} = \sum_{i_H, j_H} C_{STNorm}(i_H, j_H)$$

and $C_{ST,N}$ is the mean correlation excluding the peak area,

$$C_{ST,N} = \sum_{i,j} C_{STNorm}(i, j) - C_{ST,H}$$

Furthermore, the templates T are images of a target at specific orientations k_T. The specific orientation of the detected object is established based on the template k_S that produces the maximum SNR:

$$(k_{Smax}, k_{Tmax}, m_{Tmax}) = argmax\{SNR(k_S, k_T, m_T)\}$$

A.6 - Final Localization

Once an object has been detected, classified, and its orientation defined, the algorithm uses this information to determine the location (i_T, j_T) of the object in the source image S by choosing the peak location produced by the particular template material type and object orientation in the source image:

$$(i_T, j_T)_{max} = (i_T(k_S, k_T, m_T)_{max}, j_T(k_S, k_T, m_T)_{max})$$

Lastly, after the processing of each source image, the MAFR-TM algorithm stores whether an object has been detected, the target type, the target orientation, and the target location within the source image. The processing of the MAFR-TM algorithm is complete after each source image in a 360 [deg] survey has been processed with every one of the templates and a decision has been made.

A.7 - MAFR-TM Optimization

The computation cost of the correlation between a source image and any one template is high. Thus, we employ the Fast Fourier Transform (FFT) to compute the correlation in the wave-number domain. The source image and the templates are defined in the wave number domain and discretized. These two discrete functions are then multiplied to obtain the correlation in the wave number domain; correlation in the spatial domain of two real functions is represented by multiplication of their transforms of in the wave number domain. To reduce the wave number leakage, an optimized separable two-dimensional Kaiser window $W(p,q)$ is applied to the correlation

$$W(p,q) = \frac{I_0\left(\beta\left\{1-[(2p-I)/I]^2\right\}^{1/2}\right)}{I_o(\beta)}\frac{I_0\left(\beta\left\{1-[(2q-J)/J]^2\right\}^{1/2}\right)}{I_o(\beta)}$$

where $p = \overline{1,I}$ and $q = \overline{1,J}$ refer to the row and column indices of the window function, β controls the way the window function tapers at the edges in the spatial domain, and I_o i the zero-order modified Bessel function of the first kind. The correlation in the wave number domain is transformed back to the spatial domain using a two-dimensional inverse Fast Fourier Transform (IFFT). By using this optimized procedure, the computation count is decreased and the processing time is significantly reduced.

Appendix B: Sample Sonar Images of Various Targets in the Pool

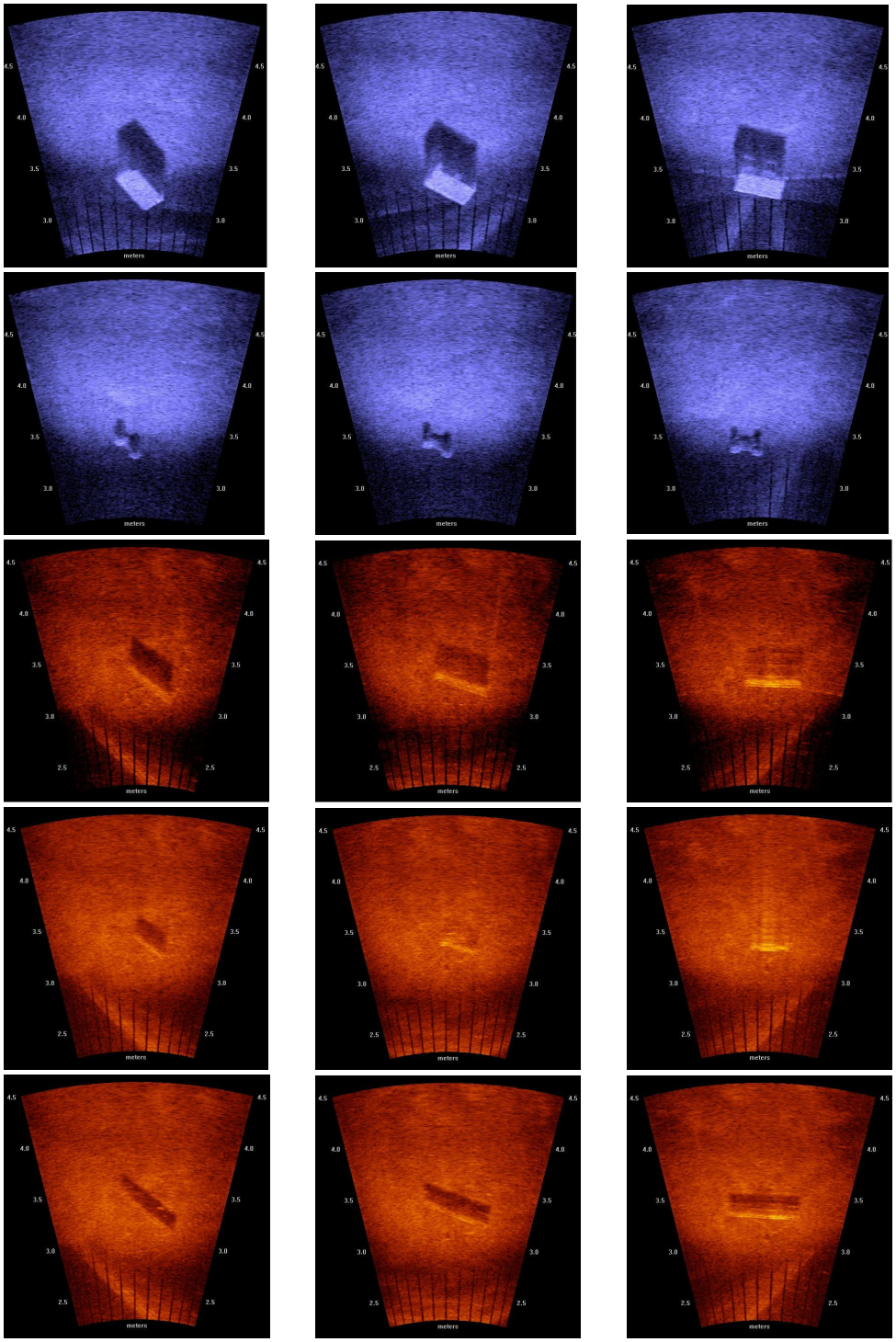
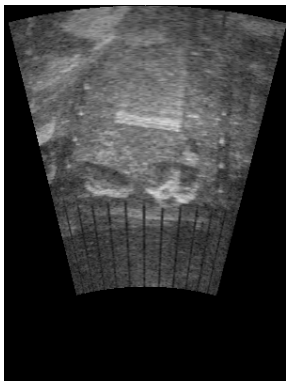
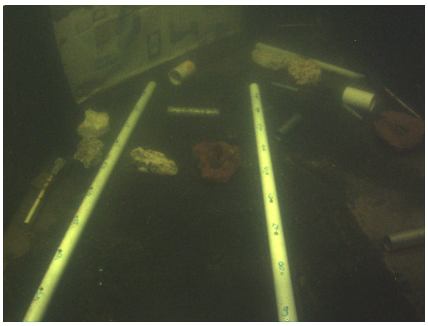
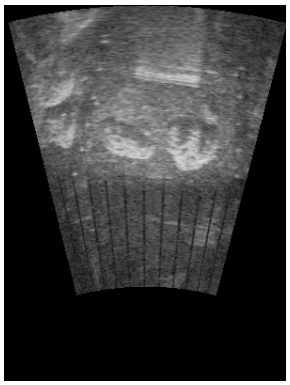
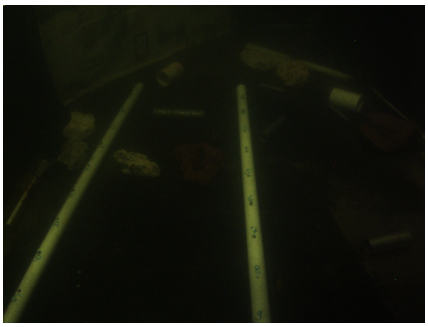
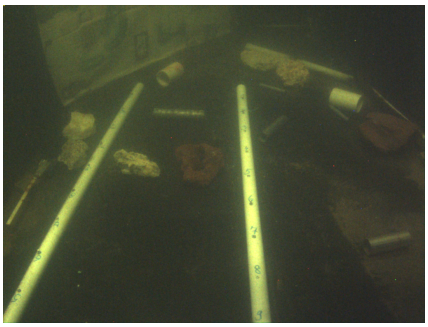
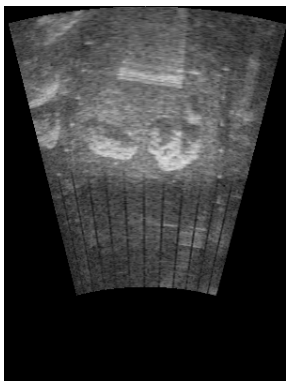
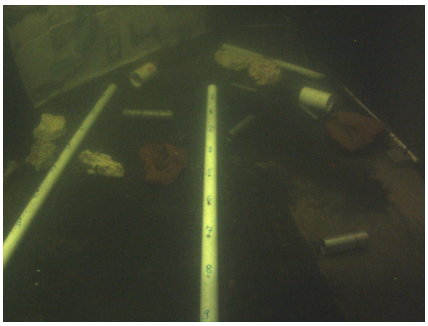
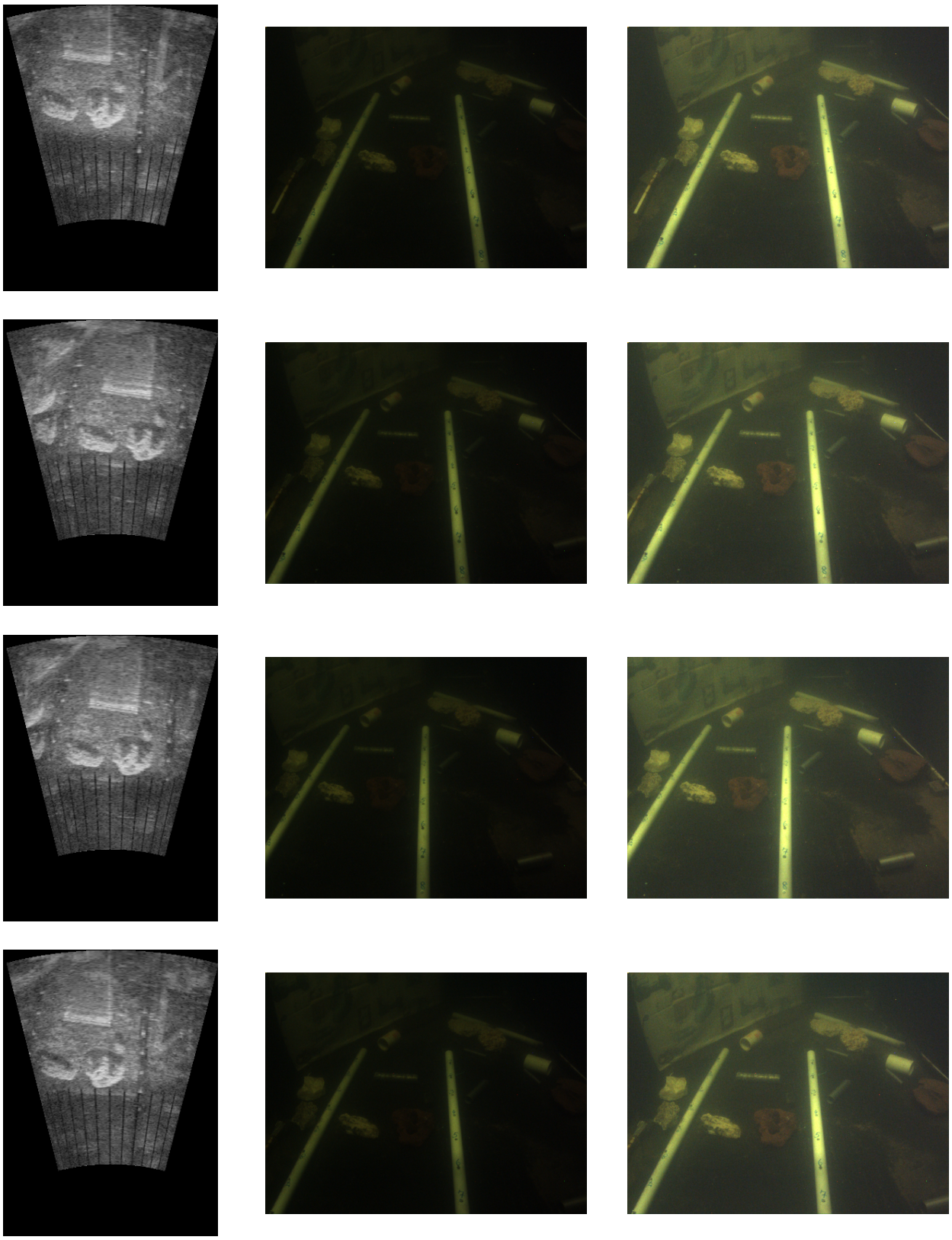
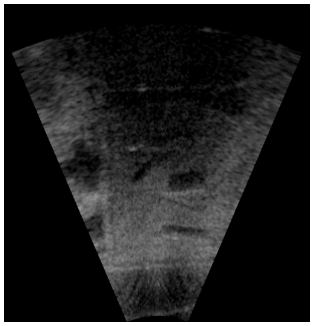
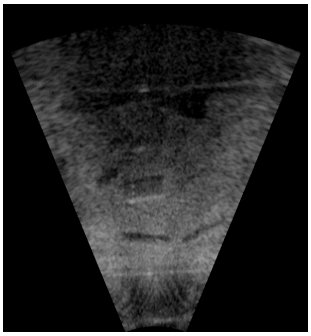
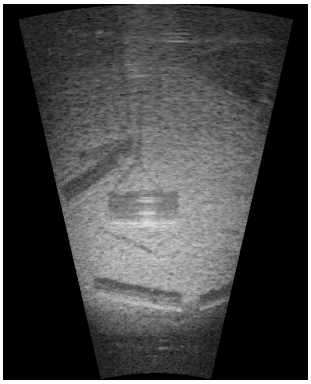
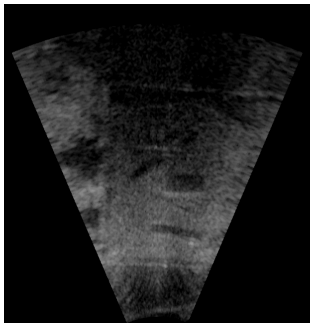
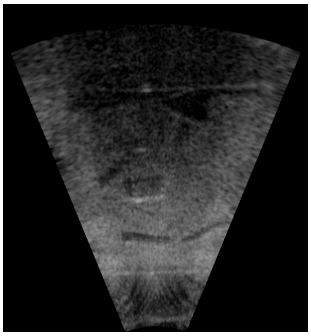
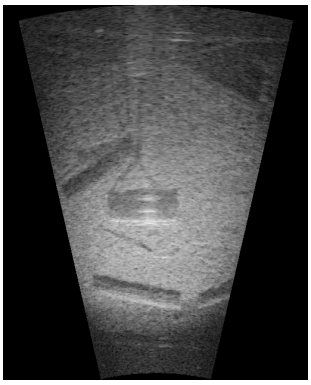
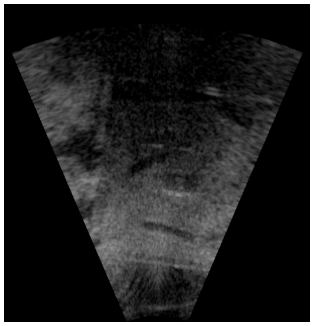
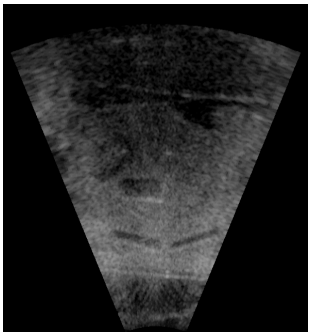
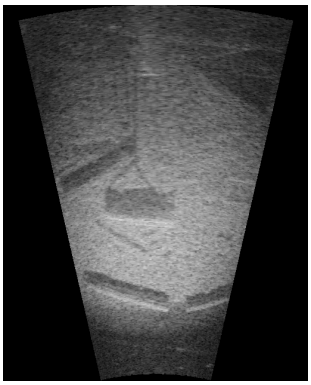
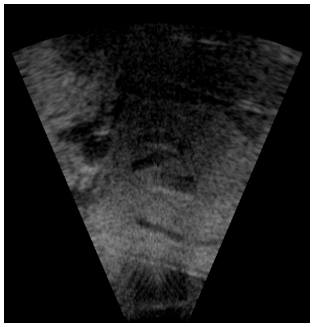
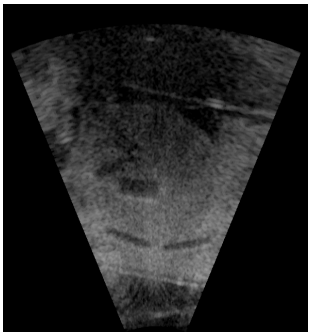
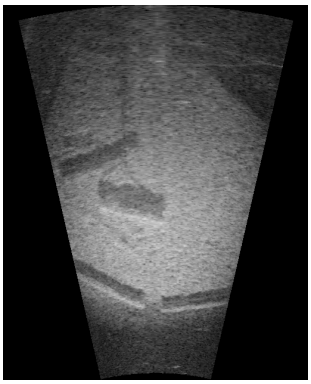
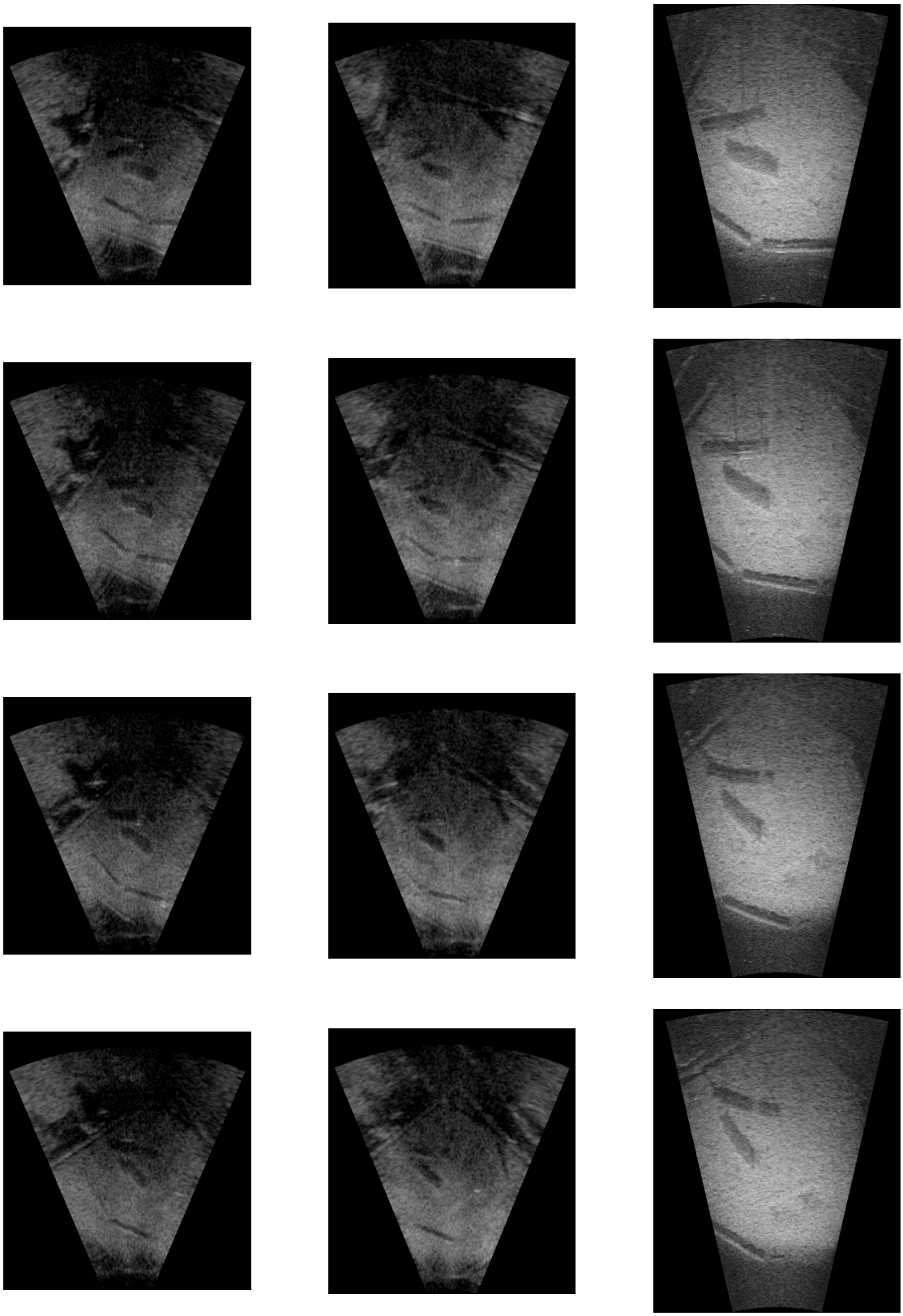
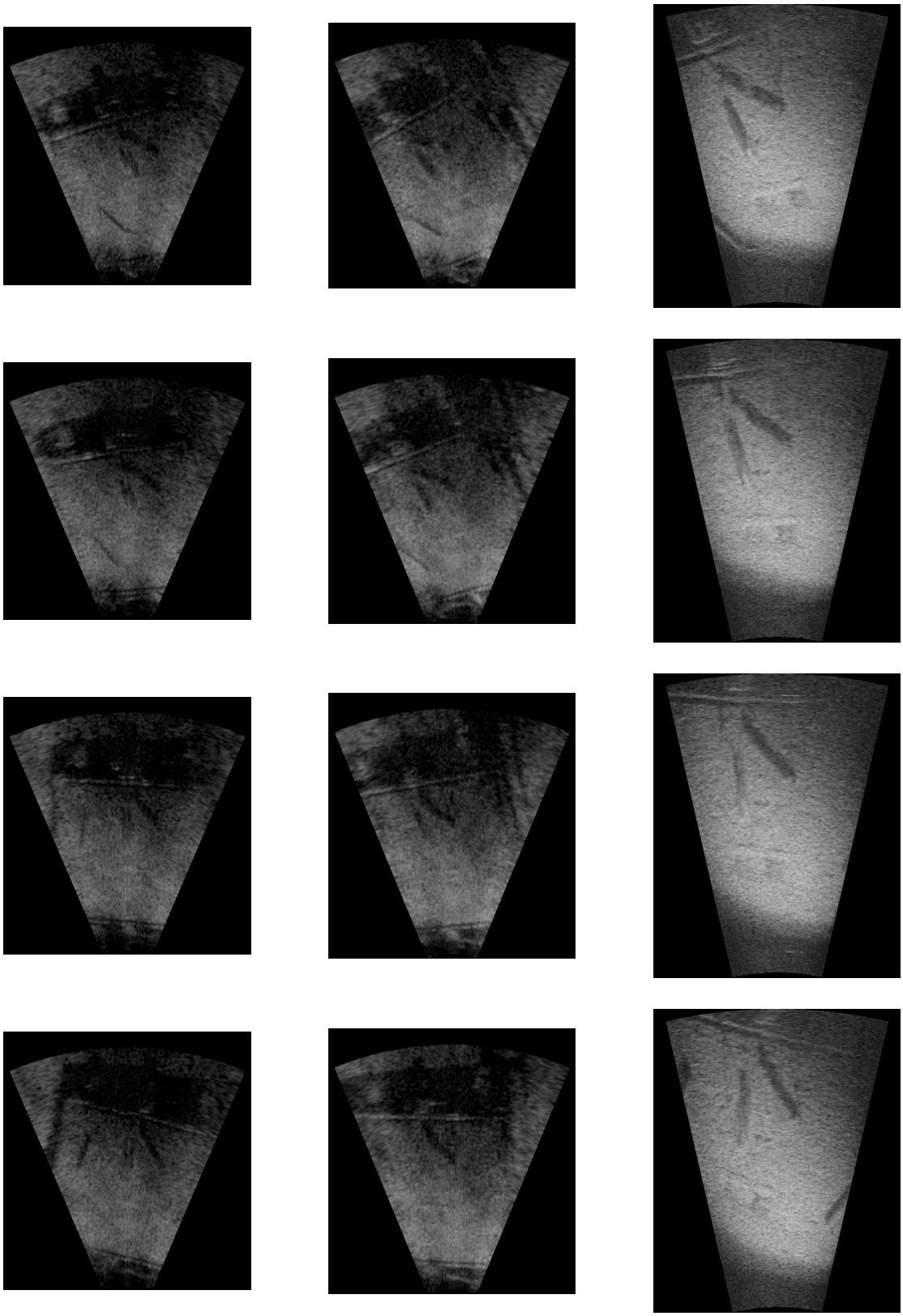
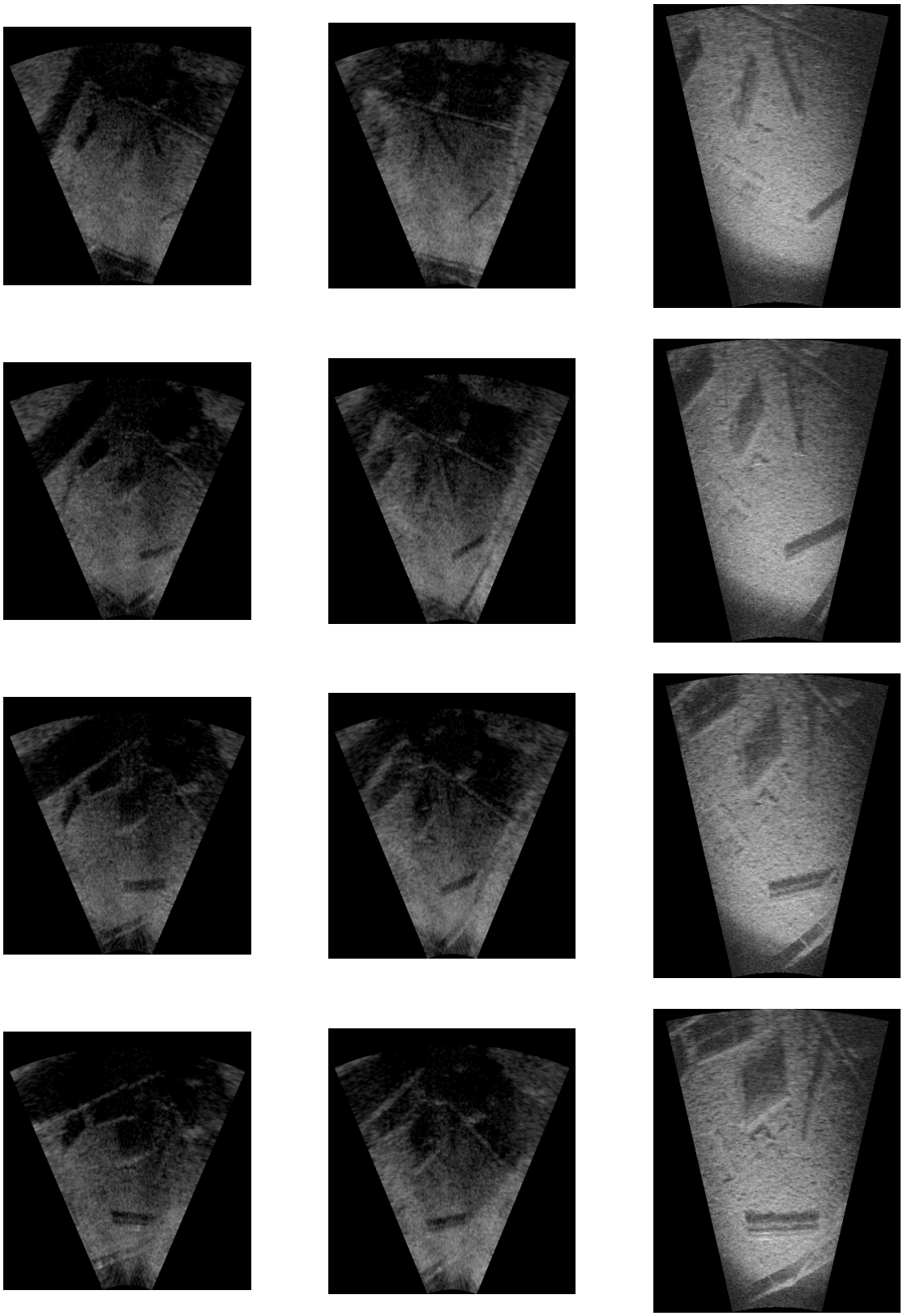
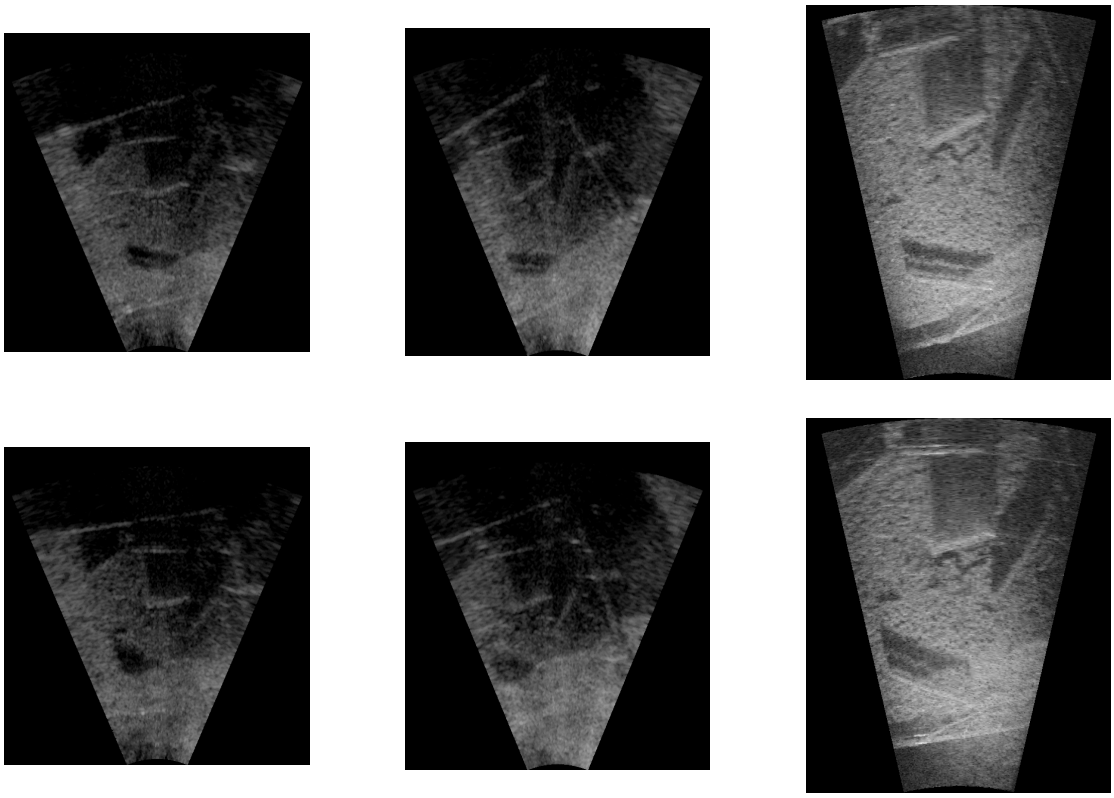


Fig. 26. DIDSON images of various targets depicted in Fig. 3 – Cinder block, dumbbell, and steel, aluminum and PVC cylinders – at 3 orientations of 320, 340 and 360 [deg] relative to the sonar.

Appendix C: Sample Images from Our Data Set

DIDSON	Optical	Optical (processed)
		
		
		



BlueView 1	BlueView 2	DIDSON