

AcTag: Opti-Acoustic Fiducial Markers for Underwater Localization and Mapping

Kalin Norman, Daniel Butterfield, and Joshua G. Mangelson

Abstract—Fiducial markers are important tools for robotic navigation and imaging, enabling accurate localization and tracking of objects in challenging environments. In this paper, we present AcTag, a new fiducial marker design for use underwater with imaging sonar and cameras, as well as a method for the detection of AcTags within acoustic images. High amounts of noise and a nonlinear projection model make it difficult to use imaging sonar in autonomous localization and mapping. In order to expand the use of imaging sonar in autonomous underwater vehicles, our marker design and detection algorithm for sonar images facilitate the identification of four unique landmarks per tag, and provide relative range and azimuth values to each landmark. We evaluate our marker design and detection algorithm with simulated and real-world sonar data, reporting on the false positive and true positive rates, as well as the estimated error for the range and azimuth estimates per landmark. We also release an open-source library for generating tag families and detecting the tags.

I. INTRODUCTION

Fiducial markers, like AprilTags, AruCo markers, and ARTags [1–3] are widely used in robotics and computer vision for automated vision-based tasks such as localization, mapping, and target tracking. Common detection algorithms are able to uniquely identify the tag and return its pose with respect to the sensor [4, 5]. In underwater environments the use of vision-based detection is limited by environmental factors such as water turbidity and reduced light levels. Those factors greatly impact the effective range and clarity of images, though visual markers still have use in marine environments [6, 7].

Acoustic sensors operate differently than cameras as the former measure returns from sound waves, while the latter measure light. Sound is able to travel for much longer distances and is relatively unaffected by turbidity, allowing sonar to operate in more diverse underwater conditions than cameras. However, standard visual fiducial markers are undetectable by acoustic sensors, and few fiducial markers designed for acoustic sensors exist.

Wang et al. and Lee et al. have designed artificial landmarks for underwater use [8, 9] that have been used within controlled environments. However, to the best of the authors’ knowledge, there are not yet any well-defined markers that can easily be used in unstructured environments or conditions. Markers that can be quickly produced, easily deployed, and detected by imaging sonar would facilitate research

This work was partially funded under Department of Navy awards N00014-21-1-2435 and N00014-21-1-2272 issued by the Office of Naval Research.

K. Norman, D. Butterfield and J. Mangelson, are at Brigham Young University in Provo, Utah.

{kalinnorman, dbutter3, joshua_mangelson}@byu.edu

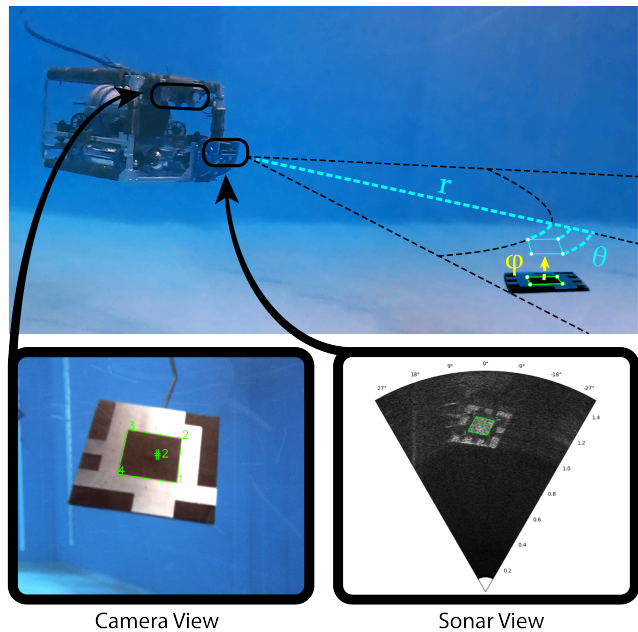


Fig. 1: Hovering Autonomous Underwater Vehicle (AUV) in water with an AcTag, a novel opti-acoustic fiducial marker that can be detected in both optical and acoustic images. Examples of acoustic and optical output of tag detection are also shown. Simultaneous detection with both camera and sonar is not always possible, so the camera and sonar views shown are from two different positions of the AUV.

in underwater localization and mapping with autonomous underwater vehicles (AUVs).

In this work we propose a marker design, that we call AcTag, that can be detected in both optical and acoustic images, as seen in Fig. 1, and an algorithm for autonomous detection in sonar imagery. The proposed algorithm is able to uniquely identify a tag under noisy conditions and provides range and azimuth measurements to the four unique corners. The design allows for the generation of large families of unique markers, similar to the aforementioned visual fiducial markers. The main contributions of this paper include:

- 1) A novel opti-acoustic fiducial marker for underwater robotic localization and mapping, with instructions for manufacturing them.
- 2) A tag family generation script modified from AprilTag 3 [1] to account for bit pattern reversal effects that occur as a result of the imaging sonar projection model.
- 3) A novel method for detection and identification of our marker in an acoustic image, resulting in range and azimuth measurements of the four uniquely identifiable corners of the tag.
- 4) An open-source library that includes both tag family

generation and tag detection functionality, found at <https://bitbucket.org/frostlab/actag>.

The rest of the paper proceeds as follows: Section II covers related works and their relation to our proposed tag design and detection method. We then cover the sonar projection model, and provide a detailed description of the tag design and construction in Section III. Tag family generation is covered in IV, followed by the details of our tag detection and decoding algorithm in Section V. An evaluation of our design and detection algorithm is presented in Section VI, using both simulated and real-world data. Lastly, Section VII outlines our conclusions, and potential future applications and developments that can be made.

II. RELATED WORK

The most closely related tag to our work is the AprilTag [1, 10, 11]. AprilTags are an open-source optical fiducial marker that have found widespread use in the area of robotics, with a robust and easy-to-use detection library. The AprilTag 3 library provides a means of generating custom tag families where the user can specify the data bits, minimum hamming distance, and layout. In 2015, Cesar et al. [6] found that AprilTags outperformed other common visual markers when used underwater in conditions where a camera could view the markers. A similar study was performed in air by Kalaitzakis et al. [4], with similar results, where AprilTags generally outperformed other tag designs in optical imagery. We frame the design of our tag families to match AprilTag, so that the existing AprilTag detection library can be used when the tags are observed using optical cameras.

There are two main contributors to the development of acoustic fiducial markers that we have identified. Wang et al. developed the ACMarker [8], which allows for 5 degree of freedom (DOF) pose estimates between the tag and the sensor. However, their pose estimate requires a fairly controlled environment, the marker is made up of large blocks of concrete, and the sensor placement relative to the marker is quite limited. Lee et al. developed another type of acoustic marker and used it in an extended Kalman filter (EKF) localization solution [9] for underwater vehicles. However, their markers are not clearly described, and rely on multiple views [12] for reliable detection.

Other opti-acoustic landmarks have been used for inter-sensor calibration, like the work presented by Yang et al. [13], where they put a large number of visual fiducial markers and magnets onto a single board. While effective for inter-sensor calibration, such boards are large, expensive, and lack the benefits that come from having more than one unique marker. Another difficulty is highlighted in the opti-acoustic calibration work by Lindzey et al [14], where they had to manually label points in the acoustic imagery. Additional efforts to create landmarks for optical and acoustic detection include that of Lagudi et al., where they used a checkerboard and metal rods wrapped in bubble wrap [15]; Raaj et al. used carefully placed buoys [16]; and Marburg et al. employed a wire-frame metal sphere [17]. All of the aforementioned methods are impressive, and effective in their own rights,

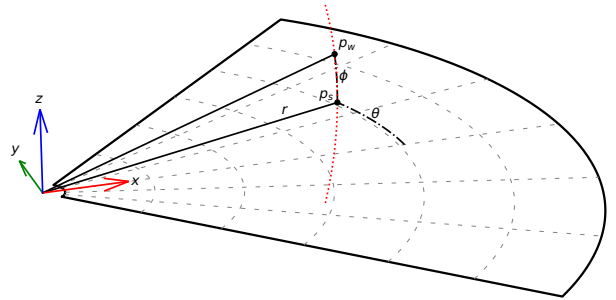


Fig. 2: Imaging sonar coordinate frame and projection model. p_w is an example of a point in the world that is observed by the sonar. Range, r , and azimuth, θ , information are preserved but the elevation, ϕ , is projected onto the range and azimuth plane. The projected point from p_w is p_s .

but lack the freedom provided by a large group of unique markers. Additionally, these solutions are often difficult to construct, especially if more than one marker is desired.

An impressive use of an opti-acoustic marker is found in the work of Negahdaripour et al. where they are able to estimate motion over time, and structure from motion [18, 19]. The features used in their work are not uniquely identifiable, and higher accuracy estimates rely on a large baseline between two views from the sensors. However, their work leverages a very simple board with holes drilled through it. Our markers are not as simple as their board, but provide much more information from a single view.

In this study we build upon these works by proposing a new fiducial marker system that combines visual and acoustic markers to enable accurate localization and mapping in underwater environments. Our system leverages the strengths of both visual and acoustic sensing, and provides a flexible, relatively low-cost, and scalable solution for underwater navigation. Detection of our markers does not rely on any manual labelling, and only a single view is necessary to obtain multiple unique landmarks.

III. PHYSICAL TAG DESIGN

A. Sonar Operation and Projection Model

Imaging sonar operates by emitting sound waves into the environment, and then measuring the sound waves that are reflected back to the sensor by the environment. Range information is estimated via time-of-flight, while bearing is measured via beamforming using an array of hydrophones to estimate the azimuth angle of return. Treating this as a spherical coordinate system, range and azimuth are preserved, while elevation is lost. Everything above or below the range-azimuth plane is projected onto the plane, along the elevation arc, as shown in Fig. 2. The resulting two-dimensional image encodes the accumulated intensity of returns at each range/azimuth cell within the field of view of the sensor [20]. Some sonars are designed to have very narrow vertical apertures, minimizing the effects of this projection. These sonars are often referred to as profiling

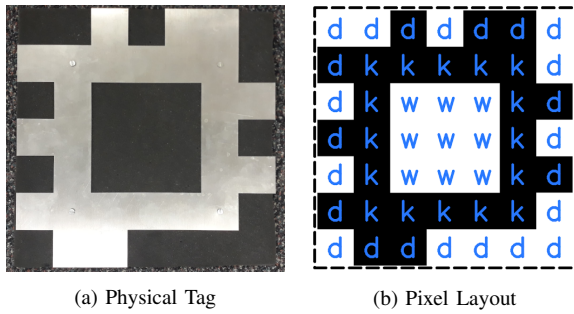


Fig. 3: AcTag construction and layout. (a) A physical AcTag. The first and third (not visible) layers are 1 mm of aluminum, of alloys 5052 or 6063, and the second layer is 12 mm of acoustic foam. (b) The layout string for a 24 data bit tag family overlaid onto an example tag. 'd' denotes a data bit, 'k' a black bit, and 'w' a white bit.

sonars and are typically very expensive. In this paper we focus on imaging sonars that have a non-negligible vertical aperture.

In this paper we represent points in the sonar's frame with a spherical coordinate system, (r, θ, ϕ) , where r is the range in meters, θ the azimuth in radians, and ϕ the elevation in radians. Given the minimum and maximum range values (r_{min} and r_{max}) in meters, vertical aperture (v) in degrees, and horizontal aperture (h) in degrees, the limits of the sonar's field of view can be described by

$$\begin{bmatrix} r_{min} & r_{max} \\ \theta_{min} & \theta_{max} \\ \phi_{min} & \phi_{max} \end{bmatrix} = \begin{bmatrix} r_{mn} & r_{mx} \\ -\frac{h}{2} \frac{\pi}{180} & \frac{h}{2} \frac{\pi}{180} \\ \frac{\pi}{2} - \frac{v}{2} \frac{\pi}{180} & \frac{\pi}{2} + \frac{v}{2} \frac{\pi}{180} \end{bmatrix}, \quad (1)$$

noting that the elevation values are centered around $\frac{\pi}{2}$.

Converting from spherical coordinates to a Cartesian coordinate system, (x, y, z) , can be done with

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r \sin \phi \cos \theta \\ r \sin \phi \sin \theta \\ r \cos \phi \end{bmatrix}. \quad (2)$$

When using sonar to identify the relative position of items in the environment, range and azimuth are preserved while elevation is lost in the projection to a sonar image.

Given a point in the sonar's frame in Cartesian coordinates, (x_w, y_w, z_w) we can project it to a point in the polar plane of the sonar, (r_s, θ_s) with the following equation

$$\begin{bmatrix} r_s \\ \theta_s \end{bmatrix} = \begin{bmatrix} \sqrt{x_w^2 + y_w^2 + z_w^2} \\ \arctan 2(y_w, x_w) \end{bmatrix} \quad (3)$$

To generate a sonar image we then convert into image coordinates, (x_s, y_s) , via

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} r_s \cos \theta_s \\ r_s \sin \theta_s \end{bmatrix}, \quad (4)$$

and bin x_s and y_s into image pixel coordinates, which we refer to as (u_s, v_s) . Usually, the top left corner of the sonar image corresponds to the furthest range and largest negative azimuth angle and rows correspond to range and columns to azimuth.

B. Tag Construction

The values encoded in a sonar image are the intensities at each range and azimuth of the return of the sound wave that was transmitted. The two main properties of materials that affect the intensity of a return when they are hit by a sound wave are 1) reflectivity, and 2) scattering coefficient. An easy assumption to make is that a very smooth surface is reflective and only generates high intensity values when the surface is perpendicular to the oncoming sound wave. On the other hand, rough surfaces tend to scatter the sound, and result in higher intensity returns to the sonar even when the surface is not perpendicular to the sound wave. This combination of smooth and rough materials can be seen in other works that have designed surfaces that are visible in imaging sonar output [8, 9, 12].

We wanted our markers to be detectable by both optical and acoustic sensors, but also to be relatively affordable and straightforward to construct. Additionally, we wanted them to be light enough to deploy and retrieve easily in real-world environments without any specialized equipment.

We found that by layering thin sheets of metal and foam, we are able to meet these requirements. For our tags we used a back plate of 1 mm thick aluminum to provide structural integrity and weight, a 12 mm thick layer of acoustic foam to scatter sound waves, and a face plate from the same material as the back plate with the tag ID pattern cut out of the metal. Four screws, with standoffs inside of the foam to maintain its shape, are used to attach the three layers together. Depending on the size of the tag, the buoyancy of the tag may need to be counteracted with additional weight.

The markers are rigid, but lightweight, and the screws through the back provide a great location to attach fishing line or other attachments that facilitate deployment and retrieval. In one use case we found that by tying small buoys to each marker we could easily identify their location within murky water. A visual of our tag design can be seen in Fig. 3(a). Within acoustic images the foam will return high intensity values, while the metal will result in low intensity returns. Optical detection is possible due to the contrast between the foam and metal. Additional details on the optical and acoustic detection are found in Section V.

When buying materials in bulk the average cost of a single AcTag is around \$25 US dollars, at the time of writing. Although production of our tags is more involved and expensive than visual fiducial markers, we believe that we've identified a means of constructing them that is straightforward, and relatively affordable for most use cases. Detailed information on the materials, costs, and assembly of AcTags can be found within our open-source library.

IV. TAG FAMILY GENERATION

In order to use multiple tags simultaneously it is important to have many unique IDs. The layout shown in Fig. 3(b) demonstrates the necessary parts of our tag families. Each tag contains a center square of white pixels, surrounded by a single layer of black pixels. The data bits then form a single pixel wide layer around the outer edge. Our tag families are

Algorithm 1: Single Image Tag Detection

Data: Raw sonar image

Result: Range and azimuth of the four tag corners

```
1  $img \triangleq$  input;  
2  $filt = \text{median\_filter}(img)$ ;  
3  $bnry = \text{adaptive\_threshold}(filt)$ ;  
4  $ctrs = \text{identify\_contours}(bnry)$ ;  
5  $quad = \text{fit\_quadrilaterals}(ctrs)$ ;  
6  $[rng, azi] = \text{identify\_tags}(quad)$ ;
```

modified versions of AprilTag 3 families such that every tag falls within these constraints.

Within this framework, tag families are specified by the number of data bits and the minimum hamming distance between tags in the family. An example tag family with 24 data bits and a hamming distance of 10 would be named: "AcTag24h10".

Given the similarities with AprilTag, we modified their tag family generation script to ensure a proper hamming distance and bit layout for AcTags. In decoding, the data bits for a tag are traversed in a clockwise order. To ensure the minimum hamming distance for optically detected AprilTags, it is necessary to check for four main offsets in rotation of the tag. In sonar images it is also necessary to check for a reversal in the order of the data bits due to mirroring. Mirroring occurs over the polar plane in the sonar projection model, such that a tag that is above the sonar can be mirrored across the plane, and a tag in the mirrored position will project onto the same location in the image. To prevent confusion, our tags are single sided, so rather than have the tags in these two positions project the exact same bit ordering, in one position the data bits will be in a clockwise ordering around the tag, while in the other position they will have a counter-clockwise ordering. As a result, an AcTag family will contain fewer tags than an AprilTag family with the same data bit and hamming distance specifications.

Our tag families also differ from AprilTag 3 in that our layout is constrained to what is described in the first paragraph of this section. While AprilTags can have data bits on the interior, the solid white interior of our tag aids in acoustic detection. The other constraint of a single layer of data bits simplifies construction, allowing the face plate to be a solid piece of metal. The one exception that can occur within our tag families is when there is a single black data bit in the corner, requiring just a little bit of extra metal to connect that data bit to the rest of the metal face plate.

V. TAG DETECTION

Camera detection of our markers is clearly described in the AprilTag papers [1, 11], and we refer the reader to their works for details on how to use their library. The only precursors to using their library with our tags is to build their detection library with the proper AcTag family and to invert the colors of the gray scale image. This inversion is necessary because the metal is lighter in color than the foam,

so the tag colors in an optical image are the opposite of those found in an acoustic image. While this could be resolved with paint, that would modify the acoustic properties of the materials and complicate tag assembly. The simplest solution we found is to invert the colors prior to optical detection.

Detection of a tag in the image from an imaging sonar is broken down into the main steps highlighted in Algorithm 1, and outputs from each step are shown in Fig. 4. Given the raw data, we treat the image as a 2D image in Euclidean space with the underlying assumption that straight lines in 3D space can be approximated by straight lines in the 2D polar image [21]. This assumption does not hold well as the tag gets close to the sonar, but the limited field of view of an imaging sonar often prevents the tag from being fully visible at close range.

In the following subsections we denote images with x and y , being inputs and outputs respectively, and their values at row i and column j as $x_{i,j}$ or $y_{i,j}$.

A. Median Filter

In order to remove some of the noise while retaining edges in the sonar image, we use a median filter. An example of the output of the median filter is shown in Fig. 4(b).

A neighborhood of pixel values around $x_{i,j}$ is defined as

$$N[x_{i,j}] = [x_{i-R,j-R}, \dots, x_{i-R,j+R}, \\ x_{i-R+1,j-R}, \dots, x_{i+R,j+R}] \quad (5)$$

where $N[x_{i,j}]$ is the neighborhood, and R is the specified radius of the neighborhood. The median filter then calculates the output values as

$$y_{i,j} = \text{median}(N[x_{i,j}]), \quad (6)$$

where $y_{i,j}$ is the output pixel value.

While simple, the filter is quite effective and has found use in other recent imaging sonar applications [8, 22].

B. Adaptive Threshold

The filtered image is then binarized through the use of an adaptive threshold, an example of which is in Fig. 4(c).

The threshold value for each pixel is set through

$$T[x_{i,j}] = \frac{\min(N[x_{i,j}]) + \max(N[x_{i,j}])}{2} - O, \quad (7)$$

where O is an offset value that can be specified by the user to raise or lower the threshold across the whole image. The output value is then determined through

$$y_{i,j} = \begin{cases} 0, & \text{if } x_{i,j} < T[x_{i,j}] \\ 1, & \text{if } x_{i,j} \geq T[x_{i,j}] \end{cases} \quad (8)$$

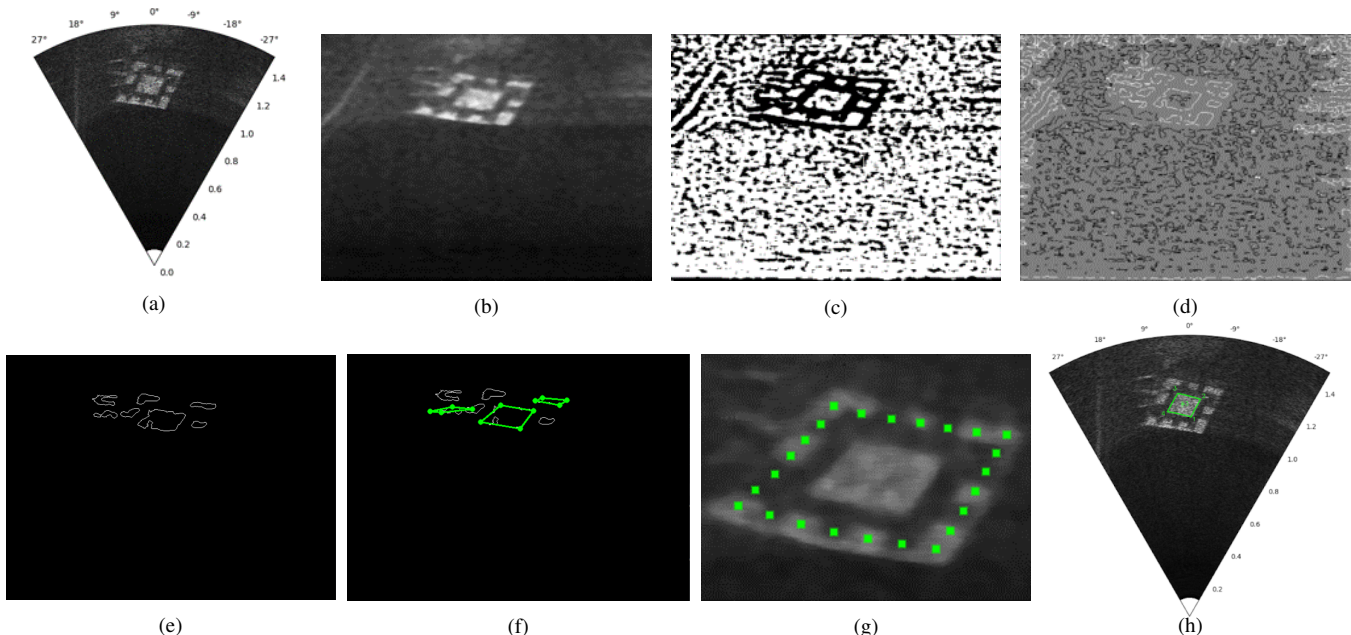


Fig. 4: AcTag sonar image detection algorithm breakdown. (a) The raw sonar image, in polar coordinates. (b) Median filter output, in Cartesian coordinates. (c) Adaptive threshold binary output. (d) All possible contours in the image. (e) Filtered contours, leveraging range and bearing from the sonar to exclude contours whose area was too large or small to be a tag. (f) Quadrilaterals that were successfully fit onto some of the contours. (g) Zoomed in view, showing sampling locations for estimated data bits around the quad to determine each data bit’s value. Note that our algorithm uses the binary image (instead of the filtered image depicted above) for sampling values. (h) Decoded tag ID and corners overlaid on the original image.

Algorithm Parameter	Value
median filter radius	4
adaptive threshold radius	8
adaptive threshold offset	1
quadrilaterals desired inlier percentage	0.85
quadrilaterals inlier distance	2.5

TABLE I: Algorithm parameters used in evaluation

C. Contour Identification

To identify contours in the image we take the binary output and use a modified version of the algorithm described by Haig et. al. in 1992 [23]. We only care about white shapes that are surrounded by black. As seen in Fig. 4(d), we are able to clearly identify both white shapes surrounded by black, and black shapes surrounded by white. Removing the black shapes surrounded by white results in far fewer contours to consider. Due to the fact that range information is returned by the sonar, additional contour rejection can be performed by calculating the estimated area of a contour. Leveraging the known range and azimuth resolutions, we can approximate the surface area that the shape occupies. If the area is obviously too small or too large to be our tag, then we can disregard the contour. The result after applying both contour rejection methods is seen in Fig. 4(e).

D. Fitting Quadrilaterals

After contour identification and early rejection, we attempt to fit a quadrilateral to each of the remaining contours. We implement a custom RANSAC [24] algorithm that attempts to fit four lines to the shape, and then determines the number of inliers and whether or not the shape defined by the intersections of those lines is a valid parallelogram. The distance for a point to be considered an inlier can be set by the user,

as well as the desired inlier percentage. Algorithm 2 outlines the main steps for fitting a quadrilateral to a single contour. If a potential quad is found to have enough inliers to be considered a good fit, the RANSAC function will terminate with a successful fit. We then refine the corner estimates by individually considering each of the four line segments formed by the corners, fitting a new line to the inliers that correspond to that line segment, and then recalculating all of the corners of the quadrilateral. Any refined quadrilateral that is also a parallelogram is then returned by the function for consideration in tag identification. An example that shows the corner locations for the quadrilaterals that were successfully fit to contours is seen in Fig. 4(f).

E. Tag Identification

Using the corners of the quadrilaterals, the known size of the tag, and the known number of data bits around the outside edge of the tag, we then identify the pixel at the center of where each data bit should be, as shown in Fig. 4(g). The binary values of the data bits are determined by sampling the binary image at those locations. The bit values are stored in a list in clockwise order and compared to the values in the look up table for the tag family. If a match is found then we obtain the tag ID and are able to pinpoint the order of the corners of the tag in the image, with Fig. 4(h) showing a successful tag detection in a sonar image. The allowable number of bit corrections is user specified and is limited by the hamming distance used in tag family generation. The minimum bit correction value is 0, and the maximum is either $(\text{hamming_dist}/2) - 1$ or $\text{floor}(\text{hamming_dist}/2)$ for even and odd values of *hamming_dist* respectively.

In the end, four unique point landmarks are clearly iden-

Algorithm 2: Fit Quad to Contour

Data: List of image coordinates in contour, c

Result: Four corners of the quadrilateral

```
1  $T \triangleq$  max attempts;
2  $N \triangleq$  inlier percentage threshold;
3  $l \triangleq$  parameters for each line;
4 while attempts <  $T$  do
5    $c_s \triangleq$  subset of contour points;
6    $n \triangleq$  number of inliers;
7    $c_s = c$ ;
8    $n = 0$ ;
9   for  $i = 0$  to 4 do
10     $pts = \text{pick\_random\_point\_group}(c_s)$ ;
11     $l[i] = \text{fit\_line\_to\_points}(pts)$ ;
12     $n = \text{determine\_inliers}(c_s, l[i]) + n$ ;
13     $c_s = \text{trim\_inliers\_from\_contour}(c)$ ;
14   $corners = \text{find\_intersections\_of\_lines}(l)$ ;
15   $n = \text{recalculate\_inliers}(corners, c)$ ;
16  if  $n/\text{len}(c) > N$  then
17     $corners = \text{refine\_corners}(corners, c)$ ;
18    if  $corners$  is parallelogram then
19      return  $corners$ 
```

tified, with observed range and azimuth values for each.

VI. EVALUATION

In this paper we focus on evaluation of acoustic detection as this is the novel contribution of our tag design. The tag fits within the definition provided by AprilTag 3 [1]. When the tag can be detected by a camera we have no reason to believe that its performance and use cases will differ from the results obtained by other researchers [6, 7].

In order to evaluate the performance of our acoustic detection algorithm we have collected both real-world and simulated data. By observing the tags with multiple different sonars at their respective frequencies we are able to identify the frequency range in which our tag is visible. With simulated data we are able to compare the range and bearing estimates for each point with the ground truth and determine approximate error distributions. Lastly, with real-world data sets we evaluate the false positive and true positive rates of our detection algorithm.

For all of our results we ran the tag detection with the parameters found in Table I.

For the false and true positive results we used a Blueprint Subsea Oculus m1200d set to a frequency of 2.1 MHz with a minimum range of 0.1 meters, a maximum range that varied between 3 and 6 meters, a horizontal aperture of 60 degrees, and vertical aperture of 12 degrees. Our simulated results were set to use the same sonar parameters as the false and true positives, with two exceptions: the maximum range was set to 5 meters, and there was no parameter for operating frequency.

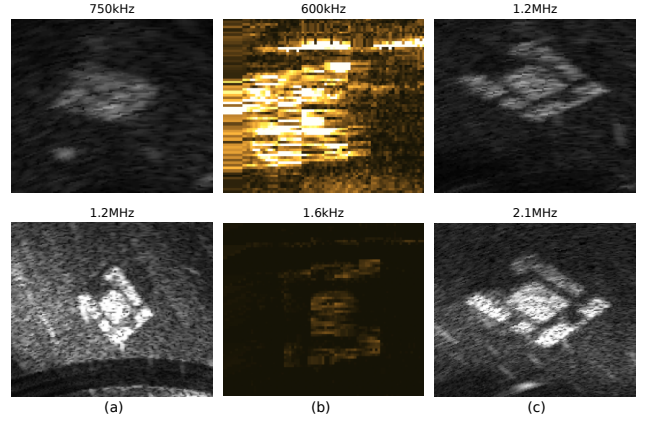


Fig. 5: AcTag viewed by three different sonars and in their respective low and high frequency modes. (a) is from a Blueprint Subsea Oculus m750d, (b) is a sidescan sonar, and (c) is a Blueprint Subsea Oculus m1200d. The tag design is difficult to view in sidescan imagery, though it still demonstrates the frequencies that can be used to detect the tag more clearly.

A. Frequency Analysis

Material properties vary with frequency, meaning that our tag will not look the same to every sonar, as they may operate over a broad frequency spectrum. We were able to test the visibility of our tag construction with a Blueprint Subsea m750d imaging sonar, a sidescan sonar, and a Blueprint Subsea m1200d imaging sonar. Each of these were equipped with a low and high frequency mode, with the lowest and highest frequencies being 600 kHz and 2.1 MHz, respectively. From the results that can be seen in Fig. 5, we determined that the tag is visible between 1.2 and 2.1 MHz, with the possibility of being visible beyond 2.1 MHz. The proposed detection algorithm in this paper does not work in sidescan imagery, where only range information is gathered and must be stitched together over time to get an image like the one shown in Fig. 5(b). However, we include sidescan output here because it was useful for testing additional frequencies not sampled by our imaging sonars.

B. Error Distributions from Simulation

Using the HoloOcean simulator [25, 26], we created a simulated tag and used a simulated AUV equipped with an imaging sonar in order to approximate error distributions. A visualization of the simulation environment can be seen in Fig. 6(a). While HoloOcean does not have a built in asset for our tags, we were able to approximate one by creating an object in the shape of a tag with occupied and unoccupied pixels, where the occupied pixels represented foam, and the unoccupied ones represented metal. The simulation parameters were tuned by visually comparing a simulated and real-world image, with a visual comparison found in Fig. 6. Given the differences between our simulated tag and a real tag, we did not feel it appropriate to evaluate the true and false positive rates within simulation. However, the simulated images reasonably approximate the noise found in real-world images, allowing us to approximate the error in the final range and azimuth results in images where a tag was detected. The error is computed by comparing the

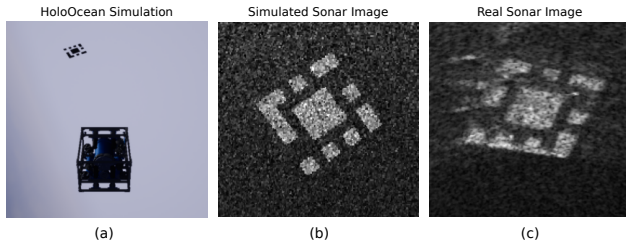


Fig. 6: Comparison of the HoloOcean simulation to real-world sonar imagery. (a) The HoloOcean environment, which consists of an AUV and an AcTag. (b) The tag within a simulated sonar image, while (c) A tag within a real-world sonar image.

algorithm’s output to the ground truth that is readily available in simulation with the known tag and sonar pose.

Results were obtained by randomly moving the sonar to a position within 5 meters of the tag, ensuring that the tag face was oriented so that it would be seen in the sonar output. Range and azimuth estimates for the four unique points on our tag were obtained from our detection algorithm, and error was computed using the ground truth range and azimuth values. Individual error distributions for the range and azimuth can be found in Fig. 7, where both distributions resemble a normal distribution with an approximate standard deviation of 1 cm for the range error and around 0.005 radians for the azimuth error. We note that the azimuth distribution is not centered at zero and is slightly skewed due to an approximation of the atan2 function that HoloOcean uses for azimuth calculations.

Given that these results were obtained from simulation, as we had no other resources to provide reliable ground truth data, these results likely fail to capture all of the error present in real-world data. However, it does give us enough information to assume that the error can be approximated as a Gaussian, and overall it should be small, on the order of centimeters for range and hundredths of a radian for azimuth.

C. False and True Positives

In order to determine the false positive rate, multiple real-world data sets from locations including: Deer Creek Reservoir in Utah, a swimming pool at Brigham Young University (BYU), and our custom vinyl-lined test tank at BYU, were composed into a single data set with 39838 total sonar images. No tags were present in any of the images, so a detection of any tag by the algorithm is a known false positive. The results can be seen in Table II, where we tested with four different tag families, each differing only in the minimum hamming distance, and therefore the number of total possible tags in each family. Additionally, we ran each family through twice, once with zero allowable bit corrections, to minimize false positives, and once with the maximum number of bit corrections allowed for that family, to demonstrate the worst case scenario.

Even in the worst case scenario, the false positive rate is below 1% for all tag families that we tested, and in the best case it is approximately 0%.

To evaluate the true positive rate, we ran the tag detection

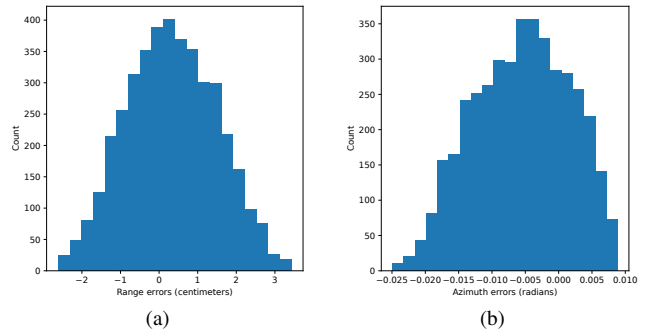


Fig. 7: Range and Azimuth error distributions from simulated data. Data shown is from 1042 tag detections, or 4168 unique points, with the errors then calculated per point. Ground truth is known in the simulation. Negative and positive errors correspond to estimated values that are smaller and larger than the true values, respectively. Only data for correct tag detections is considered.

Bit Cor.	Tag Family	Tags in Family	False Detections	False Pos. Rate
1	AcTag24h4	6261	235	0.5898891 %
2	AcTag24h6	453	247	0.6200110 %
3	AcTag24h8	45	236	0.5923992 %
4	AcTag24h10	7	281	0.7053567 %
0	AcTag24h4	6261	8	0.0200813 %
0	AcTag24h8	45	2	0.0050203 %
0	AcTag24h8	45	0	0.0 %
0	AcTag24h10	7	0	0.0 %

TABLE II: False positive evaluation where tag detection was run on a data set of 39838 images, each of which did not contain any fiducial markers. The data set was composed of images from a tank, a recreational pool, and Deer Creek reservoir in Utah. In the first four rows, the maximum number of allowable bit corrections was set for each tag family used. The last four rows correspond to no bit corrections allowed.

on a different set of data, collected at the BYU pool, that consisted of 21228 images. Most, if not all, of the images have a single tag in them, as it was carefully collected to keep the tag within the field of view of the sonar. The tag was 1 ft. square and was typically between 1 and 4 meters away from the sonar. The results are in Table III. For these results we only considered it to be a true detection if the algorithm found the correct tag ID within the allowed number of bit corrections. Erroneous detections are when either the wrong tag was identified, or an additional tag was identified in the image. Both the count of the erroneous detections and the true positive rates are reported in the table.

A trade-off exists between false and true positives, especially in the use of bit corrections. Allowing the maximum number of bit corrections will result in a higher number of true positive detections, while also increasing the number of erroneous detections, and the false positive rate. When bit corrections are used, the hamming distance for the family greatly impacts the true positive rate, while having little effect on the false positive rate. Smaller family sizes with greater hamming distance between the tags in the family result in the greatest chance of correct tag identification, while retaining similar false positive rates to other families with more tags and a smaller hamming distance.

We note that the rates reported herein depend on the tag families used, the tag size, as well as the parameters set within the tag detection algorithm.

Bit Cor.	Tag Family	Tags in Family	Err. Det.	True Det.	True Pos. Rate
1	AcTag24h4	6261	1128	4132	19.465 %
2	AcTag24h6	453	646	6801	32.038 %
3	AcTag24h8	45	547	8943	42.128 %
4	AcTag24h10	7	799	10545	49.675 %
0	AcTag24h4	6261	39	1568	7.387 %
0	AcTag24h6	453	3	1568	7.387 %
0	AcTag24h8	45	0	1568	7.387 %
0	AcTag24h10	7	1	1568	7.387 %

TABLE III: True positive evaluation where tag detection was run on a data set of 21228 images from a recreational pool, the majority of which contained a marker within them. In the first four rows, the maximum number of allowable bit corrections was set for each tag family used. The last four rows correspond to no bit corrections allowed.

VII. CONCLUSION

In conclusion, we have presented a new fiducial marker designed to work with both cameras and imaging sonar. The marker is constructed with a solid aluminum back, a thin acoustic foam layer in the middle, and an aluminum front with a unique tag ID cutout. Tag families can be generated to include a desired number of data bits, with a required hamming distance between each tag within the family.

We have described an algorithm for detecting the tag in sonar images, which includes noise reduction through a median filter, an adaptive threshold to binarize the image, identification of contours, a custom RANSAC implementation to fit quadrilaterals to the contours, and lastly the identification of tags within the specified tag family. Our algorithm identifies the tag ID, orientation, and range and azimuth for the four uniquely identifiable corners of the tag.

Our contributions include the development of a relatively affordable and reasonably easy to make fiducial marker, a custom tag family generator, the ability to identify four point landmarks with known range and azimuth values relative to the sonar's pose, and an open-source library that contains everything needed for tag family generation and detection.

We believe that our work has significant potential in robotic localization and mapping in underwater environments. Our open-source library can be used by other researchers and practitioners in the field. Through this, we hope that our work will stimulate further research in this area.

A future development would be to extract a full 6 DOF relative pose estimate, as is currently possible for visual fiducial markers.

REFERENCES

- [1] M. Krogus, A. Haggemiller, and E. Olson, "Flexible layouts for fiducial tags," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019, pp. 1898–1903.
- [2] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, Jun. 2014.
- [3] M. Fiala, "ARTag, a fiducial marker system using digital techniques," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2005, pp. 590–596 vol. 2.
- [4] K. Michail, B. Cain, S. Carroll, A. Anand, W. Camden, and V. Nikolaos, "Fiducial markers for pose estimation," *Journal of Intelligent & Robotic Systems*, vol. 101, no. 4, 2021.
- [5] J. G. Mangelson, R. W. Wolcott, P. Ozog, and R. M. Eustice, "Robust visual fiducials for skin-to-skin relative ship pose estimation," in *Proceedings of IEEE OCEANS MTS - Monterey*, Sep. 2016, pp. 1–8.
- [6] D. B. dos Santos Cesar, C. Gaudig, M. Fritsche, M. A. dos Reis, and F. Kirchner, "An evaluation of artificial fiducial markers in underwater environments," in *Proceedings of IEEE OCEANS - Genova*, May 2015, pp. 1–6.
- [7] E. Westman and M. Kaess, "Underwater AprilTag SLAM and calibration for high precision robot localization," Oct. 2018.
- [8] Y. Wang, Y. Ji, D. Liu, Y. Tamura, H. Tsuchiya, A. Yamashita, and H. Asama, "ACMarker: Acoustic Camera-Based Fiducial Marker System in Underwater Environment," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5018–5025, Oct. 2020.
- [9] Y. Lee, J. Choi, and H.-T. Choi, "Experimental results on EKF-based underwater localization algorithm using artificial landmark and imaging sonar," in *Proceedings of IEEE OCEANS - St. John's*, Sep. 2014, pp. 1–3.
- [10] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of IEEE International Conference on Robotics and Automation*, May 2011, pp. 3400–3407.
- [11] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2016, pp. 4193–4198.
- [12] Y. Lee, J. Choi, N. Ko, and H.-T. Choi, "Probability-Based Recognition Framework for Underwater Landmarks Using Sonar Images," *Sensors*, vol. 17, no. 9, p. 1953, Aug. 2017.
- [13] D. Yang, B. He, M. Zhu, and J. Liu, "An Extrinsic Calibration Method With Closed-Form Solution for Underwater Opti-Acoustic Imaging System," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6828–6842, Sep. 2020.
- [14] L. Lindzey and A. Marburg, "Extrinsic calibration between an optical camera and an imaging sonar," in *Proceedings of IEEE OCEANS - San Diego*, Sep. 2021, pp. 1–8.
- [15] A. Lagudi, G. Bianco, M. Muzzupappa, and F. Bruno, "An alignment method for the integration of underwater 3D data captured by a stereovision system and an acoustic camera," *Sensors*, vol. 16, no. 4, 2016.
- [16] Y. Raaj, A. John, and T. Jin, "3D object localization using forward looking sonar (FLS) and optical camera via particle filter based calibration and fusion," in *Proceedings of IEEE OCEANS MTS - Monterey*, Sep. 2016, pp. 1–10.
- [17] A. Marburg and A. Stewart, "Extrinsic calibration of an RGB camera to a 3D imaging sonar," in *Proceedings of IEEE OCEANS MTS - Washington*, Oct. 2015, pp. 1–6.
- [18] S. Negahdaripour and A. Taatian, "3-D motion and structure estimation for arbitrary scenes from 2-D optical and sonar video," in *Proceedings of IEEE OCEANS - Kobe*, Apr. 2008, pp. 1–8.
- [19] —, "3-D motion estimation by integrating visual cues in 2-D multi-modal opti-acoustic stereo sequences," *Computer Vision and Image Understanding*, vol. 114, pp. 928–941, Aug. 2010.
- [20] S. Negahdaripour, "Visual motion ambiguities of a plane in 2-D FS sonar motion sequences," *Computer Vision and Image Understanding*, vol. 116, no. 6, pp. 754–764, Jun. 2012.
- [21] N. T. Mai, H. Woo, Y. Ji, Y. Tamura, A. Yamashita, and H. Asama, "3D reconstruction of line features using multi-view acoustic images in underwater environment," in *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Nov. 2017, pp. 312–317.
- [22] J. Li, Z. Tang, Z. Wang, J. Huang, and C. Wang, "Application of artificial landmark in sonar detection," *Measurement Science and Technology*, vol. 33, no. 8, p. 085407, Aug. 2022.
- [23] T. Haig, Y. Attikiouzel, and M. Alder, "Border following: new definition gives improved borders," *IEE Proceedings I (Communications, Speech and Vision)*, vol. 139, no. 2, pp. 206–211, Apr. 1992.
- [24] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
- [25] E. Potokar, S. Ashford, M. Kaess, and J. G. Mangelson, "HoloOcean: An Underwater Robotics Simulator," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, May 2022, pp. 3040–3046.
- [26] E. Potokar, K. Lay, K. Norman, D. Benham, T. B. Neilsen, M. Kaess, and J. G. Mangelson, "HoloOcean: Realistic Sonar Simulation," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2022, pp. 8450–8456.