Estimation of Target Orientation in SAS images

Herman Midelfart and Øivind Midtgaard

***Abstract*—The orientation of an object is important for classification of underwater objects in sonar images. It is needed both for robust feature extraction and for efficient template matching. Hence, we examine several methods for estimating the orientation. Two common mine shapes, cylinders and wedges, are considered, and the best method for each shape is determined. This evaluation is conducted on both simulated and measured SAS images.**

***Index Terms*—Image orientation analysis, object recognition, image classification, synthetic aperture sonar**

I. Introduction

Classification of objects in sonar images is often performed by dividing the object response into a highlight and a shadow segment. Geometric, statistical and texture features are then extracted from the segments and put into a classifier that assigns a class to the object. Another common approach to classification is template matching. This approach compares a sonar image of an object to a set of templates and assigns the object to the class of the most similar template. A template is a reference image and is typically created by a simulator from 3D models of possible targets.

In both approaches, it is advantageous to know the orientation of the object in the horizontal plane. To compute geometric features such as object length and width, one needs the orientation to position a coordinate system onto the object. If this coordinate system has incorrect orientation, the estimates of the length and the width will also be inaccurate. In template matching, one needs to test a set of templates with the target portrayed at different orientation angles. One may, for example, create a template for every 10^{th} degree, yielding a total of 36 templates for a target. Such a set of templates must be created for every target class unless the number of templates can be reduced due to target symmetry. Hence, with a large number of targets the computational effort may be considerable. This is especially problematic if the templates are not prepared in advance, but rather created on the fly during the matching process. If the orientation is known, on the other hand, it is sufficient to compare the image to only one template for each target class. The knowledge of the orientation may thus reduce the computational effort. Moreover, it may increase the specificity of the template matching process since fewer templates are matched to an object so that the risk of a false incidental match is reduced. If the templates are also created on the fly, the orientation in the templates can be made to exactly match the object in the image, which may increase the classification sensitivity.

This study is part of FFI's development of an Automatic Target Recognition (ATR) system for seafloor mines in Synthetic Aperture Sonar (SAS) imagery. In this paper, we examine how the object orientation in the horizontal plane can be estimated for two common mine shapes: Cylinders and wedges. We apply several different methods such as moment estimation, the Radon transform, the (generalized) Hough transform, most frequent pair estimation and minimum bounding box estimation and compare their performances. The evaluation is conducted on both simulated and measured SAS images.

The paper is organized as follows: Sec. II presents the mine shapes and the methods for determining their orientation. The results are reported in Sec. III. We finish with conclusions in Sec. IV.

Herman Midelfart and Øivind Midtgaard are with the Norwegian Defence Research Establishment (FFI), P. O. Box 25, NO-2027 Kjeller, Norway (Phone: +47 63 80 70 00; fax: +47 63 80 71 15; e-mail: {Herman.Midelfart, Oivind.Midtgaard}@ffi.no).

II. Methods

A. Mine shapes

We consider two different mine shapes in this study: Cylinders and wedges.

Many mine types are shaped like a cylinder. Recognizing the orientation of such a shape is thus of high practical importance. These mines may vary in length, but have usually quite similar radius to enable deployment through the torpedo tubes of submarines. The length is typically much larger than the radius. This gives them an almost rectangular highlight in the sonar image and makes it easy to separate the short ends from the main sides. An important property of this object is that the main sides are parallel to the major axis of the object so that it is sufficient to estimate the orientation of the main sides to determine the orientation of the object.

The wedge is a more complex shape. It has several properties that make estimation of the orientation more difficult. The bottom is shaped like a trapezoid with the height of the trapezoid parallel to major axis of the object. Its major axis thus is not parallel to any of the sides. Hence, the orientation cannot be estimate directly from an edge in the image. One could possibly add a small term to the estimated orientation to compensate for the difference between the main side and the major axis. However, the length and the width are of almost identical size. So even if one knows the orientation of one side, one still has to determine which side of the wedge is actually considered. This cannot be done without considering the whole object or at least several of its sides.

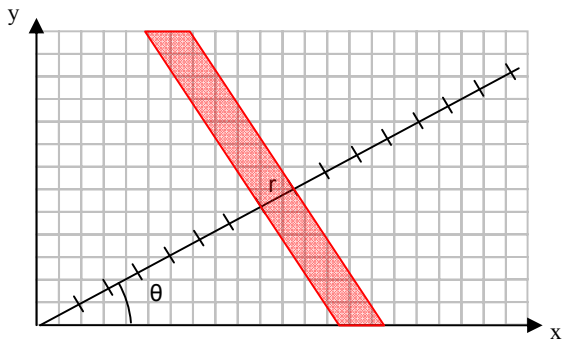


Fig. 1. The Radon Transform: The values of the pixels in the red area are aggregated into $g(r, \theta)$.

B. *Methods for estimating orientation of cylinders*

There are several methods that can be used to estimate the orientation in the horizontal plane (i.e., the rotation angle around depth axis) of a cylinder. All of these methods except for the Radon transform require that the highlight of the object has been identified in the image. To achieve this we use an adaptive, region-growing segmentation algorithm related to [1]. It starts with a region consisting of a single pixel and grows the region recursively by examining the adjacent pixels. A pixel is added to the region if its intensity is above a threshold. This threshold depends on intensity range in the image and the number of neighbor pixels that have already been included in region.

Radon Transform. One estimate of the orientation can be obtained with the Radon transform [2-4]. This method projects all pixels in an image onto a line and is computed for every possible line. A discrete form of the Radon transform $g(r, \theta)$ can be defined as

$$g(r,\theta) = \sum_x \sum_y f(x,y)\delta(x\cos\theta + y\sin\theta - r) \qquad (1)$$

where r denotes a distance from the origin along a line, and the angle θ defines the slope of the line. The image value at pixel coordinate (x, y) is given by $f(x, y)$, and the function $\delta(x)$ is 1 if $\lfloor |x| + 0.5 \rfloor = 0$ and 0 otherwise. The transform has the effect that contributions of pixels lying on a line perpendicular to line with angle θ' and at distance r' from origin is aggregated into the accumulator $g(r', \theta')$ (see Fig. 1). If a contact is oriented along this perpendicular line, it will give a large contribution to $g(r', \theta')$ as most of its pixels will be aggregated into $g(r', \theta')$. If it has a different orientation, fewer pixels will contribute to $g(r', \theta')$, and the value will be lower. The orientation of the object can thus be estimated by finding the angle θ that has the highest value in $g(r, \theta)$.

Some modifications to the image are necessary to obtain a robust estimate with this transform. We normalize and log-transform an image before we apply the Radon transform so that highlight pixels has a value larger than 0, shadow pixels have a value smaller than 0 and background pixels have value around 0. Moreover, all pixels below a certain threshold (> 0) are removed. Shadow pixels are consequently not considered in this analysis.

Hough Transform. This method is very similar to the Radon transform. The main difference is that the Hough transform [5;6] must be applied to a binary image where the image $f(x,y)$ has a value of 1 for the pixels corresponding to the object highlight and 0 otherwise. To achieve this we apply the segmentation algorithm which creates a binary image. The Hough transform is then applied to this binary image, and the orientation is estimated in same manner as the Radon transform.

Moments. The axis of least moment inertia (see e.g., [4;7]) is the axis that minimizes

$$I(\theta) = \sum_{(x,y)\in R} \left((x-\bar{x})\sin\theta - (y-\bar{y})\cos\theta\right)^2 \qquad (2)$$

Here, θ denotes the angle between the axis and the x-axis. R is a set of coordinates corresponding to the pixels in highlight region. The center of the region is defined by the means $\bar{x} = \frac{1}{|R|}\sum_{[x,y]\in R} x$ and $\bar{y} = \frac{1}{|R|}\sum_{[x,y]\in R} y$. $I(\theta)$ is the sum of the squared distance between a pixel and the center projected onto the perpendicular axis (to the one given by θ). Minimizing $I(\theta)$ with respect to θ yields the solution

$$\theta = \frac{1}{2}\tan^{-1}\frac{2\mu_{11}}{\mu_{20}-\mu_{02}} \qquad (3)$$

which we use as an estimate of the cylinder orientation. The moments μ_{pq} are defined as

$$\mu_{pq} = \sum_{[x,y]\in R} (x-\bar{x})^p (y-\bar{y})^q \qquad (4)$$

Note that a similar estimate can also be obtained with Principal Component Analysis by estimating the orientation of the first principle component.

Most Frequent Pair. Neulist and Armburster [8] introduce a method for estimating the orientation by computing the orientation of every possible pair in a point cloud. The pair orientations are then clustered, and the centroid of the largest cluster is used as an estimate. We apply this method to estimate the orientation of the highlight. But rather than cluster the pair orientations, we follow Palm [9] and estimate the orientation from the mode of the probability density computed from the pair orientations. More precisely, we begin by segmenting the image and extracting the boundary of the highlight segment. The orientation of each pair of points on this boundary is computed with respect to the x-axis. The most frequent orientation is then used as an estimate of the object orientation.

Minimum Bounding Box. This is a very simple method that is applied to the highlight segment. A bounding box is the smallest rectangle that encloses the segment. The method searches for the bounding box by examining a multitude of rotation angles. For each rotation angle it constructs a rectangle that encloses the segment and is oriented according to the angle. The rectangle with the least area is the bounding

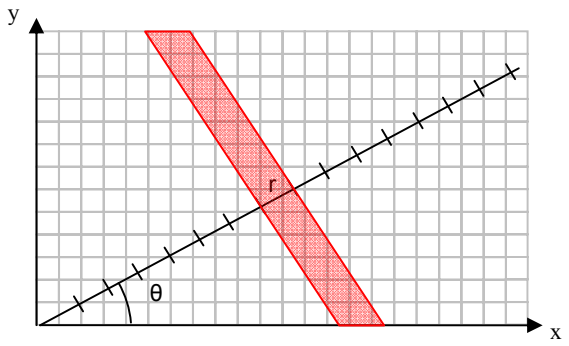


Fig. 1. The Radon Transform: The values of the pixels in the red area are aggregated into $g(r, \theta)$.

B. *Methods for estimating orientation of cylinders*

There are several methods that can be used to estimate the orientation in the horizontal plane (i.e., the rotation angle around depth axis) of a cylinder. All of these methods except for the Radon transform require that the highlight of the object has been identified in the image. To achieve this we use an adaptive, region-growing segmentation algorithm related to [1]. It starts with a region consisting of a single pixel and grows the region recursively by examining the adjacent pixels. A pixel is added to the region if its intensity is above a threshold. This threshold depends on intensity range in the image and the number of neighbor pixels that have already been included in region.

Radon Transform. One estimate of the orientation can be obtained with the Radon transform [2-4]. This method projects all pixels in an image onto a line and is computed for every possible line. A discrete form of the Radon transform $g(r, \theta)$ can be defined as

$$g(r,\theta) = \sum_x \sum_y f(x,y)\delta(x\cos\theta + y\sin\theta - r) \qquad (1)$$

where r denotes a distance from the origin along a line, and the angle θ defines the slope of the line. The image value at pixel coordinate (x, y) is given by $f(x, y)$, and the function $\delta(x)$ is 1 if $\lfloor |x| + 0.5 \rfloor = 0$ and 0 otherwise. The transform has the effect that contributions of pixels lying on a line perpendicular to line with angle θ' and at distance r' from origin is aggregated into the accumulator $g(r', \theta')$ (see Fig. 1). If a contact is oriented along this perpendicular line, it will give a large contribution to $g(r', \theta')$ as most of its pixels will be aggregated into $g(r', \theta')$. If it has a different orientation, fewer pixels will contribute to $g(r', \theta')$, and the value will be lower. The orientation of the object can thus be estimated by finding the angle θ that has the highest value in $g(r, \theta)$.

Some modifications to the image are necessary to obtain a robust estimate with this transform. We normalize and log-transform an image before we apply the Radon transform so that highlight pixels has a value larger than 0, shadow pixels have a value smaller than 0 and background pixels have value around 0. Moreover, all pixels below a certain threshold (> 0) are removed. Shadow pixels are consequently not considered in this analysis.

Hough Transform. This method is very similar to the Radon transform. The main difference is that the Hough transform [5;6] must be applied to a binary image where the image $f(x,y)$ has a value of 1 for the pixels corresponding to the object highlight and 0 otherwise. To achieve this we apply the segmentation algorithm which creates a binary image. The Hough transform is then applied to this binary image, and the orientation is estimated in same manner as the Radon transform.

Moments. The axis of least moment inertia (see e.g., [4;7]) is the axis that minimizes

$$I(\theta) = \sum_{(x,y)\in R} \left((x-\bar{x})\sin\theta - (y-\bar{y})\cos\theta\right)^2 \qquad (2)$$

Here, θ denotes the angle between the axis and the x-axis. R is a set of coordinates corresponding to the pixels in highlight region. The center of the region is defined by the means $\bar{x} = \frac{1}{|R|}\sum_{[x,y]\in R} x$ and $\bar{y} = \frac{1}{|R|}\sum_{[x,y]\in R} y$. $I(\theta)$ is the sum of the squared distance between a pixel and the center projected onto the perpendicular axis (to the one given by θ). Minimizing $I(\theta)$ with respect to θ yields the solution

$$\theta = \frac{1}{2}\tan^{-1}\frac{2\mu_{11}}{\mu_{20}-\mu_{02}} \qquad (3)$$

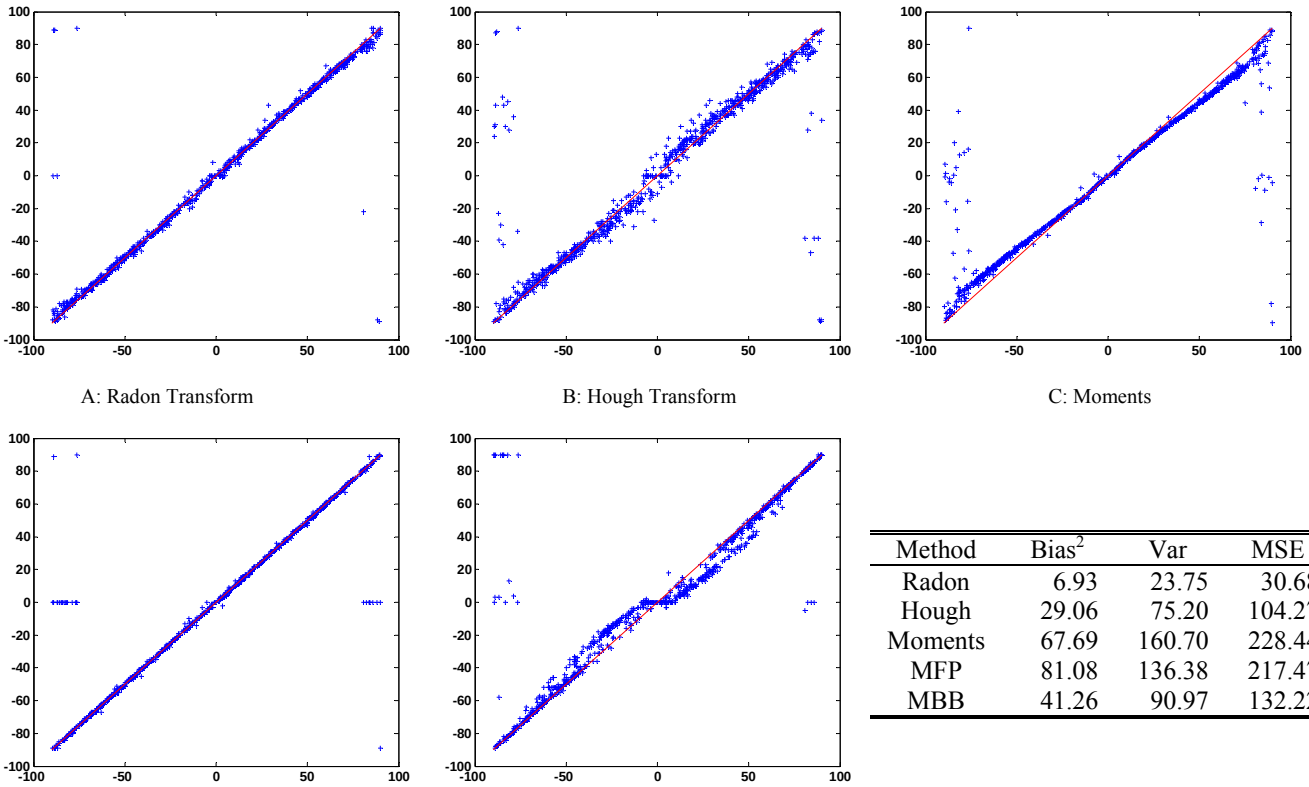
which we use as an estimate of the cylinder orientation. The moments μ_{pq} are defined as

$$\mu_{pq} = \sum_{[x,y]\in R} (x-\bar{x})^p (y-\bar{y})^q \qquad (4)$$

Note that a similar estimate can also be obtained with Principal Component Analysis by estimating the orientation of the first principle component.

Most Frequent Pair. Neulist and Armburster [8] introduce a method for estimating the orientation by computing the orientation of every possible pair in a point cloud. The pair orientations are then clustered, and the centroid of the largest cluster is used as an estimate. We apply this method to estimate the orientation of the highlight. But rather than cluster the pair orientations, we follow Palm [9] and estimate the orientation from the mode of the probability density computed from the pair orientations. More precisely, we begin by segmenting the image and extracting the boundary of the highlight segment. The orientation of each pair of points on this boundary is computed with respect to the x-axis. The most frequent orientation is then used as an estimate of the object orientation.

Minimum Bounding Box. This is a very simple method that is applied to the highlight segment. A bounding box is the smallest rectangle that encloses the segment. The method searches for the bounding box by examining a multitude of rotation angles. For each rotation angle it constructs a rectangle that encloses the segment and is oriented according to the angle. The rectangle with the least area is the bounding



A: Radon Transform

B: Hough Transform

C: Moments

D:. Most Frequent Pair (MFP)

E: Minimum Bounding Box (MBB)

Method	$Bias^2$	Var	MSE
Radon	6.93	23.75	30.68
Hough	29.06	75.20	104.27
Moments	67.69	160.70	228.44
MFP	81.08	136.38	217.47
MBB	41.26	90.97	132.22

Fig. 2. The orientation estimation for the cylinder on the simulated data set: The actual orientation is given on horizontal axis of the scatter plots, and the estimated orientation is shown on the vertical axis. The blue crosses correspond to images and should ideally lie along red line. The mean square error (MSE), the variance and the squared bias are given for each method in the table.

The Hough transforms had more problems with short end images since it was applied to the segmented image. The variance of its estimates was also larger. This was also a result of the segmentation. The highlight segment was often almost shaped like a thin rectangle, and since the diagonal of this rectangle was slightly longer than any of the sides, it could contribute more to the accumulator. Thus, estimate became slightly distorted. The Radon transform was probably less sensitive to this problem since it also used the intensity in the image, and the intensity provided additional information about the orientation.

We applied also the Hough transform to the boundary of the highlight segment and ignored the rest of the segment so that such diagonals would not give a large contribution. In this case, we obtained an estimate with lower variance (Result not shown).

The axis of least moment inertia method produced also incorrect estimates for the short end facing cylinders. However, this was not only a consequence of the incorrect segmentation as this method was biased. At 0 degrees, the estimate was accurate, but as the aspect angle approached either -90 or 90 degrees the difference between the actual and estimate angle increased.

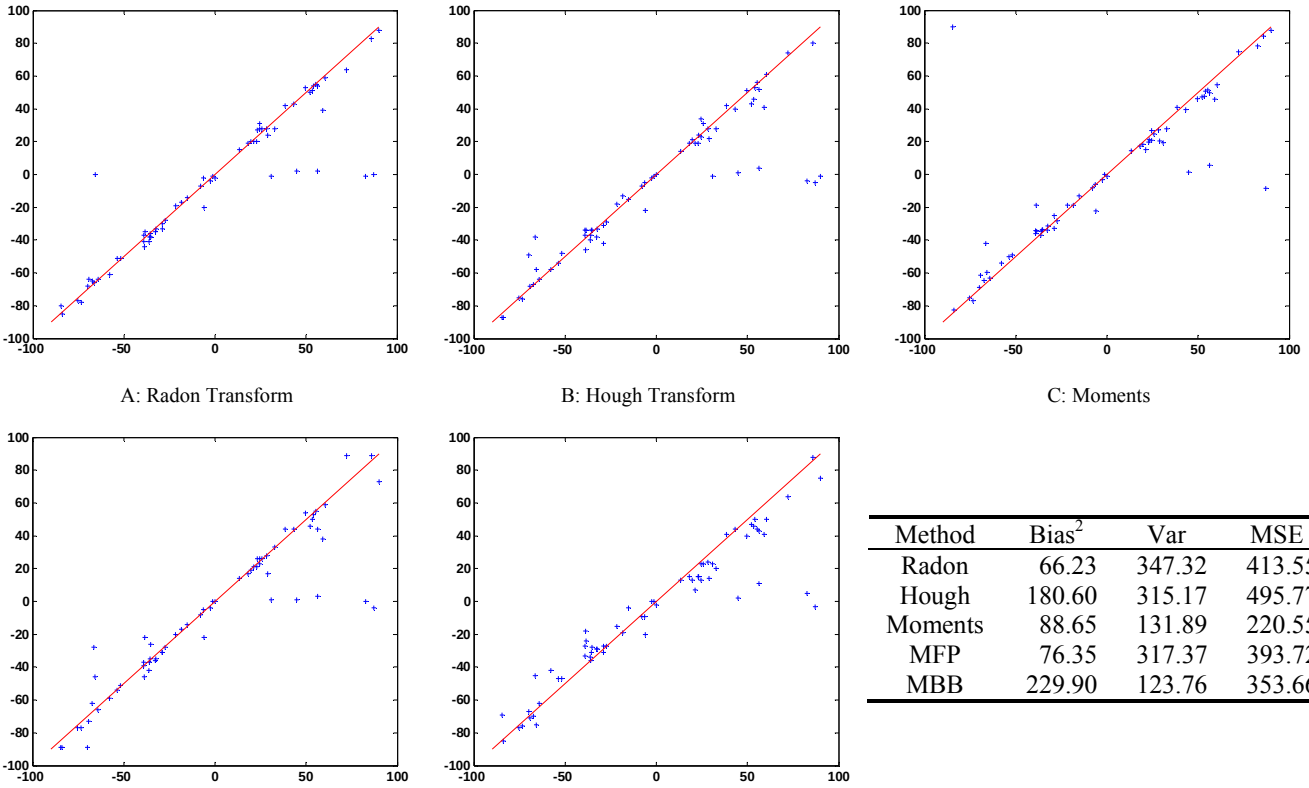
The most frequent pair method provided a very accurate estimate. However, it had more problems with the short end cylinders than the other segment-based methods. Many of the cylinders with orientation around -90 and 90 degrees received an estimate of zero degrees.

The minimum bounding box had some systematic errors. It estimated the orientation accurately at 0 degrees. However, for angles slightly above 0 and up to approximately 45 degrees it underestimated the orientation by 10 degrees. The angles slightly below 0 and down to approximately -45 degrees were similarly overestimated by 10 degrees. Short end facing cylinders were also incorrectly estimated.

The results from the real data are shown in Fig. 3. All of the methods estimated the orientation accurately for most of the cylinders. However, some mistakes were made, and the same mistakes were made by several methods. Six cylinders were given Radon transform estimates that were far from the actual value. These six suffered from image artifacts of various degree. In three images the object had a thick line parallel to the along-track axis in addition to the real highlight. This line gave the dominant contribution to the accumulator so that the orientation was estimated for it and not from the real highlight. Two of the other images were difficult cases with the cylinders oriented close to 90 degrees and the highlight stretched out in the along-track direction. They were thus given an estimate of 0 degrees.

The performance of the Hough transform and the most frequent pair method were quite similar. They made the same mistakes as the Radon transform except for one image. For a couple of images their estimates were less accurate than those of the Radon transform.

The best results were obtained with the moments and the most frequent pair methods. They failed only on 3 or 4 of the



A: Radon Transform

B: Hough Transform

C: Moments

D: Most Frequent Pair (MFP)

E: Minimum Bounding Box (MBB)

Method	Bias^2	Var	MSE
Radon	66.23	347.32	413.55
Hough	180.60	315.17	495.77
Moments	88.65	131.89	220.55
MFP	76.35	317.37	393.72
MBB	229.90	123.76	353.66

Fig. 3. The orientation estimation for the cylinder on the real data set: The actual orientation is given on horizontal axis of the scatter plots, and the estimated orientation is shown on the vertical axis. The blue crosses correspond to images and should ideally lie along red line. The mean square error (MSE), the variance and the squared bias are given for each method in the table.

images that were incorrectly estimated by the Radon transform. Hence, they may be more robust on distorted images, but there is not much evidence in favor of such a conclusion. These methods were also biased on the simulated data. If one looks closely at the scatter plots, one might perceive the same effects as witnessed on the simulated data. However, they are not very clear since the data set was limited.

B. Estimation of wedge orientation

To determine the orientation of the wedge we used the GHT together with the 2D model shape shown in Fig. 4e. We examined several versions of the GHT. Initially, we tried the standard version described in Sec II.C. In order to apply this method we needed the boundary of the image. This was acquired by segmenting the image and tracing the boundary of the highlight segment. The gradient was also computed for the boundary pixels. The estimates that we obtained with this method were inaccurate as shown in Fig. 4a. One reason is that the edges of the boundary were uneven so it was difficult to obtain good estimates of the gradient. Hence, the lookup in the R-table failed so that many relevant image pixels did not contribute to the accumulator

Thus, we tried to run the algorithm with R-table lookup disabled so that contributions were made to the accumulator for every pixel on the extracted boundary. The result was much better in this case as shown in Fig. 4b. The mean squared error (MSE) was reduced with 42 %. The procedure took slightly more time, but was still fast since the extracted boundary contained relatively few pixels.

We applied also the GHT to the full highlight segment so every pixel in the segment contributed to the accumulator. The R-table was not used in this case either. This computation took much more time. However, the result (see Fig. 4c) was also better with a 57 % reduction in MSE from the standard GHT.

The transform was finally applied to the image (still without the R-table). This requires considerably more execution time. In order to reduce this effort, the transform was only used in a window around the highlight. The result was even better with 76 % drop in MSE from the standard GHT. However, there were still some errors as shown in Fig. 4d. Most of these appeared in two areas. Wedges between 0 and 20 degrees were often given an estimate around -150 to -140 degrees, and wedges between 160 and 180 degrees received estimates between -30 and -50 degrees. Closer examination of the accumulator revealed that it had two peaks in these cases, and the wrong peak had been chosen. This occurred since one side gave stronger backscatter than the rest of the wedge body. Consequently, this side dominated the accumulator. Since two edges of the model shape could be matched to this side, two peaks appeared in the accumulator.

IV. CONCLUSION

We have examined several methods for estimating the orientation of a target in SAS images. For cylinder objects, we

box, and its rotation angle is used as an estimate of the true orientation. More information can be found in e.g., [4].

C. Method for estimating orientation of wedges

The methods described above provide only an estimate in the interval between -90 and 90 degrees. Moreover, they require that the object is elongated. These limitations do not fit well with the wedge shape. Its length and the width are most similar, and the sides are not parallel with the major axis as mentioned above. The wedge is not symmetric with respect to its minor axis (in the horizontal plane) like the cylinder and requires a method that can produce an estimate in an interval of 360 degrees.

Generalized Hough Transform. One applicable method is the Generalized Hough Transform (GHT) that can be used on arbitrary shapes to determine their position, scale and orientation. To achieve this it extends the Hough Transform accumulator with more dimensions. Typically, it may have dimensions such as x and y for position, σ for scale and θ for orientation. We assume, however, that the object has a fixed size so that we will not consider the scale.

The GHT works in a same way as the Hough transform, but instead of aggregating the pixels on a line in the image into a single pixel in the accumulator, it aggregates the pixels corresponding to a model shape into a single pixel. Formally, we may define the shape with a parametric curve $\vec{s}(t) = x(t)\vec{\imath} + y(t)\vec{k}$. This shape may be translated and rotated so the shape is represented in the image by

$$\vec{\omega}(t, \vec{r}, \theta) = \vec{r} + R(\theta)\vec{s}(t) \qquad (5)$$

where $\vec{\omega}(t, \vec{r}, \theta)$ denotes an image coordinate, $\vec{r}$ defines a translation of the shape, and $R(\theta)$ is a rotation matrix for a rotation angle θ. The vector $\vec{r}$ defines the location of the shape in the image. It also defines the point in the accumulator that receives the pixel values corresponding to the shape. So, a position $\vec{\omega} = [x, y]^T$ in the image corresponds to a point $\vec{r}$ in the accumulator for a given orientation θ and parameter t. The accumulator $g(\vec{r}, \theta)$ may be defined

$$g(\vec{r}, \theta) = \sum_x \sum_y \sum_t f(x, y)\delta\left([x, y]^T - \vec{r} - R(\theta)\vec{s}(t)\right) \qquad (6)$$

Where $\delta([a, b]^T) = \delta(a)\delta(b)$. The image *f*(*x*, *y*) is usually binary and contains only the boundary of the object.

A comparison of the expressions in (1) and (6), reveals the GHT requires more computational effort than the Radon and the Hough transform. For a given pixel (*x*, *y*) in the image and an orientation θ, the parameter r can be determined uniquely in the Hough transform, but we obtain multiple positions $\vec{r}$ in the GHT since the shape is parameterized on t. Hence, a pixel in the image is considered many more times in the GHT.

However, it is not necessary to consider all of the positions since there is only one shape point $\vec{s}(t)$ that corresponds to an image pixel if the image and the shape actually match. The GHT introduces therefore an additional constraint to exclude shape points that cannot possibly match an image pixel. This is done by requiring that a pixel in the image contributes only if it has the same gradient as the corresponding shape point. To perform this matching efficiently the GHT uses an R-table. This is an array organized according to the gradient of the shape. Each entry in the array holds the shape points with that particular gradient. When a pixel in the image is matched to the shape, the algorithm looks up the entry with the same gradient as the image pixel and adds only the values corresponding to these shape points to the accumulator.

The generalized Hough transform was introduced Ballard [10]. A good introduction can be found in [6]. It is also related to template matching.

III. Results

The methods were tested on both real and synthetic images. The real data set was acquired with the HISAS 1030 sonar onboard a HUGIN 1000 AUV. It consisted of images from two different locations outside of Horten in Norway. A cylinder exercise mine was deployed as a test object in both locations, and the data set consisted primarily of images of this object. However, examination of the SAS images revealed that one of the areas contains a lot of scattered debris, including many cylinder shaped objects. These cylinder shaped objects were also included in the data set. The real data set contained a total of 65 images. No wedge shapes were included.

We applied the sonar simulator SIGMAS to create simulated images of the cylinder and the wedge. More details on SIGMAS can be found in [11;12]. A total of 815 images were created of the cylinder, and 867 images were created of the wedge. The altitude (5 to 40m), the range (0 to 200m), the burial level (0 to 0.1m), the aspect angle (-180 to 180 degrees), the rotation angle around the along-track axis (-5 to 5 degrees), and the rotation angle around the range axis (-5 to 5 degrees) were varied during the simulation as specified in the parentheses. The altitude and the range were restricted so that grazing angle was less than 60 degrees.

A. Estimation of cylinder orientation

The results that we obtained on the simulated images of the cylinder are shown in Fig. 2. The aspect angle was measured clockwise from the along-track axis to the major axis of the cylinder. Hence, one main sides of the cylinder was facing the sonar at zero degrees, and one of short ends was pointing to the sonar at -90 and 90 degrees.

The mean squared error (MSE) and the corresponding bias and variance of each method are given in the table of Fig. 2. In this computation, the difference between the actual and the estimated orientation was computed modulo 180 degrees.

Most of the methods struggled when the short end of the cylinder was facing the sonar. The reason was that the backscatter from the top of the cylinder was often weak and almost indiscernible from the seabed. The segmentation algorithm created, in this case, a highlight just for the short end facing the sonar, but not for the top side. As this highlight was shaped like a rectangle with the longest side in the along-track direction, the estimate became zero degrees.

One exception was the Radon transform. It was applied directly to the image so its estimate did not depend on the segmentation algorithm. This method performed better or just as well as any other method for other the orientation angles. Thus, it had generally the best performance.