Robust Object Classification in Underwater Sidescan Sonar Images by Using Reliability-Aware Fusion of Shadow Features

Naveen Kumar, *Student Member, IEEE*, Urbashi Mitra, *Fellow, IEEE*, and Shrikanth S. Narayanan, *Fellow, IEEE*

***Abstract*—Detecting and classifying objects in sidescan sonar images is an important underwater application with relevance to naval transportation and defense. Properties of the imaging modality, in this case, often introduce large intraclass variabilities reducing the discriminative power of any classification algorithm and limiting the possibilities of improving classification accuracy by advances in pattern recognition only. In this work, we investigate the role of an ancillary feature set computed on object shadows and propose a scheme for exploiting this useful, but variedly reliable information for object classification. A mean-shift-clustering-based segmentation technique is used for isolating highlight and shadow segments from the images. We show the results of reliability-aware fusion of features computed on highlight and shadows on three different data sets of sidescan sonar images, to illustrate under what conditions such information might be useful.**

***Index Terms*—Object classification, reliability-aware fusion, shadow segmentation, sidescan sonar, Zernike moments.**

I. Introduction

OBJECT classification tasks are often domain sensitive requiring that the design of algorithms or choice of features be done on a case-by-case application basis. Detecting and classifying objects underwater is one such application domain important for naval transportation, security, and defense. Different sensing modalities are used for obtaining the signal information in this case, with sidescan sonar imaging [1], [2] being a key one among them.

There are a number of challenges for object detection and classification in underwater scenarios because of the constraints of the acoustical medium, the wide variability and heterogeneity in the underwater environment, and the objects of interest therein. In addition, they include other constraints related to costs associated with acquiring, transmitting, and processing data underwater. Hence, the data available for object classification are often not ideal, requiring robust methods for improving classification accuracy. In this work, we consider the case of underwater object classification from sidescan sonar images.

Among the different modalities available for imaging underwater, sonar imaging is often preferred over others. One reason is its ability to function in low visibility conditions unlike optical imaging methods that require additional light sources [3], [4]. In addition, low power consumption and low cost are important factors that make sonar imaging module an essential component in most autonomous underwater vehicles (AUVs). On the other hand, sonar images often undergo certain irreversible transformations due to reasons ranging from hardness/texture of the object to characteristics of the medium, or even the angle of viewing. In our current target application, large areas of the seabed are scanned at once using sidescan sonars, thereby posing greater challenges due to the difficulty in differentiating objects of interest from the sea bed (Fig. 1). Nevertheless, sidescan sonar imagery is invaluable to underwater applications like mine–countermine operation where speed is an important factor [1], [5]–[8]. An accurate and reliable recognition of a mine or other dangerous objects on the seabed could often lead to a speedy neutralization of the threat.

Since classical supervised pattern classification approaches do not adequately perform on these images, there have been attempts to use additional knowledge specific to sidescan sonars. For example, there has been ample work on extracting and computing features based on the object's shadow [9]. A variety of machine learning techniques have also been employed to improve the adaptability of the algorithms to unseen features in the test data. These include methods such as active learning [10] and classifier fusion [11]. But in spite of such advanced methods, the false alarm rate is generally deemed to be higher than acceptable for various tasks. This might still be acceptable in other applications like mine detection as long the false negatives are controlled. Thus, detection is typically followed by a second detection/classification stage, where the detected object is further classified based on its type or attribute, e.g., as a mine or a nonmine. Some of these works have tried to push the bar even higher by trying to further classify the detected object according to its shape [1]. In these settings, supplementary information such as those from object shadows could help better discriminate between different object shapes. However, the appearance of a sonar shadow in the image typically depends on a number of factors, and while the use of this additional information has been shown to be useful, it has not been systematized. In particular, the information available from the object and its

Manuscript received January 09, 2014; revised June 28, 2014; accepted July 21, 2014. Date of publication September 03, 2014; date of current version July 10, 2015. This work was supported by the U.S. Office of Naval Research (ONR) under Grant N00014-09-1-0700.

Associate Editor: J. Tory Cobb.

The authors are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: komathnk@usc.edu; ubli@usc.edu; shri@sipi.usc.edu).

Digital Object Identifier 10.1109/JOE.2014.2344971

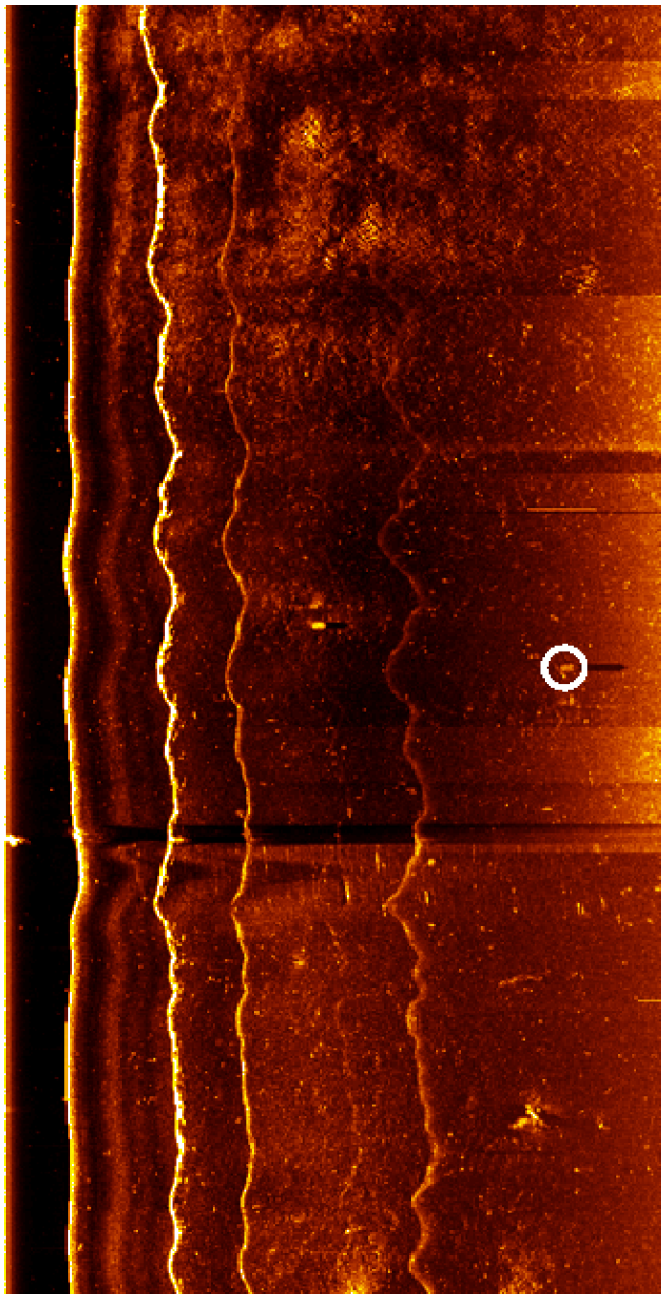


Fig. 1. A sample object (circled) in a partial sidescan sonar image with its shadow to the right. (A colormap has been added to the original grayscale image for ease of viewing.)

shadow for classification are usually not equally reliable. In this paper, we consider a data-driven reliability-aware approach for object classification in sidescan sonar images. We demonstrate its effectiveness on two data sets.

We employ an algorithm based on mean-shift clustering (MSC) to first segment the shadow and object from the sonar images. Zernike magnitude features are then used to encode the object and shadow shapes for classification. Importantly, we also propose a novel Bayesian network model to take into account the potential uncertainty associated with the shadow features. Object classification is performed by doing statistical inference on this model. In this paper, we choose a fully parametric mixture model for simplicity. More specifically, we assume that the shadow features are generated by a mixture of two distributions one of which depends on the class label (reliable) and the other is class independent (unreliable).

In fact, this approach is partly similar to the idea of garbage modeling commonly used in automatic speech recognition tasks [12], [13], where these models are used to identify out of vocabulary (OOV) words in dialog systems [14] or keyword spotting (KWS) [15] systems. Relatedly, the machine learning approach of boosting [16] uses a similar notion of weighting samples differently during training to reduce the bias in supervised learning. Boosting proceeds by correcting itself such that previously "misclassified" training samples can be correctly classified by the ensemble of weak learners.

Such methods assume features from all training samples to be equally reliable. In contrast, in this paper, we weight the samples to include or exclude them from the "garbage model" that tries to model the outlier distribution. Samples with a low reliability weight are more likely to have been generated by the "garbage model" and *vice versa*. We model this data characteristic by a Bayesian network and show how to perform the maximum likelihood (ML) parameter estimation and inference. Finally, we test the performance of our algorithm on different underwater object data sets and show that the inferred reliability values in each case match intuition (Table IV).

The paper is organized as follows. Section II introduces the object classification problem and the challenges associated with it. Section III describes the strategy used for segmentation of these objects, and Section IV presents details of feature representation including a brief review of the Zernike moments. Section V provides insight into the problems associated with direct fusion of shadow information into the classification process. Section VI describes the graphical model used for reliability-aware fusion of shadow features. We also describe the expectation–maximization (EM) algorithm used for estimation of parameters in an ML sense and some relevant techniques for inferring variables of interest. Section VII presents experimental results, while Section VIII considers the question of generalizability and model complexity in general. Section IX offers final insights into the model and future ideas.

II. Problem and Data Set Description

To enable this empirically grounded study, we rely on realistic data sets for developing and testing the proposed classification schemes. In particular, three data sets have been used in this work: the NATO Undersea Research Center (NURC, La Spezia, Italy) data set collected by the Defence Research and Development Canada—Atlantic (DRDC Atlantic, Dartmouth, NS, Canada) and NURC [17], the Naval Surface Warfare Center (NSWC, Washington, DC, USA) scrubbed image data set, and the SSPS multiresolution sonar imagery data set collected at NSWC Panama City Division (Panama City, FL, USA) comprising over 1000 images in all (Table I). These data sets contain sidescan sonar imagery in the form of 8-b grayscale images, each containing one or more synthetic mine-like objects. Each object is approximately 10–20 pixels in width and can belong to multiple classes, based on their shape (Fig. 2). Objects in the NURC data set can be from any of the seven classes described as cone, cylinder, junk, rock, sphere, wedding cake, or wedge, while objects in the NSWC or SSPS data sets are labeled

TABLE I
NUMBER OF SAMPLES AND CLASSES FOR EACH DATABASE

Database	#Samples	#Classes
NSWC	296	4
SPSS	442	4
NURC	1038	7

as belonging to classes A through D. This work deals with the problem of classifying the objects in these data sets, assuming that the object has already been localized.

A. Data Set Characteristics

While the object might usually be clearly visible as a bright highlight because of the strong reflection of sonar waves, substantial variations in object appearance can occur as a result of the seabed environments around the object. In addition, different angles of viewing the object can make the task confusing even to the expert human eye.

Thus, one class of popular techniques used for this problem compares the shapes of the object shadows against expected shapes, generated through simulations of their 3-D templates [18]. However, the data sets of interest in our work provide no additional information about the object shapes, ruling out the possibility of using such methods that use expert knowledge of expected object properties. Only information about object location and class is provided in these data sets, thus specifying the scope of the classification scenario of interest in this work.

In addition to variabilities within a database, each data set also has its own characteristics depending on the environment or the sensor used to collect the data. Such differences often make it difficult to generalize any particular algorithm considerably affecting experiment design. As an example, notice how the echo/highlight is a dominant feature of images from the NSWC and NURC databases (Fig. 3). However, in the SSPS data set, highlights are a minor feature compared to shadows that are more informative about the object shape. Similarly, we find that images in the NSWC or NURC databases often lack a noticeable shadow. These observations underscore the importance of seeking an appropriate data characterization that is cognizant of the domain. Thus, we would like our approach to be agnostic to factors such as the sensor or environment by being able to adapt to the particular domain.

B. Contributions of This Paper

In previous works such as [1], it is assumed that shadow and highlight information are two different ways of observing the same object. As a result, an implicit assumption is made about the information obtained from objects and shadows being parallel representations and hence equally reliable. We propose a data-driven model that accounts for differences in reliability of the information available from these two modalities. Our model makes an independent missing at random (MAR) assumption for the unreliable feature set using a garbage model. We present results on each of the above data sets and show that the reliability scores obtained match our intuition of reliability for both objects and shadows.

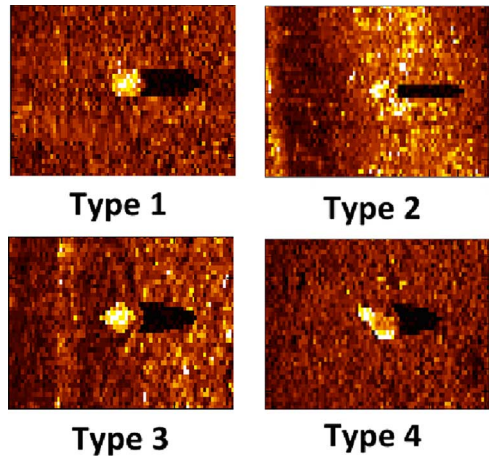


Fig. 2. Four types of object in the NSWC sidescan sonar database.

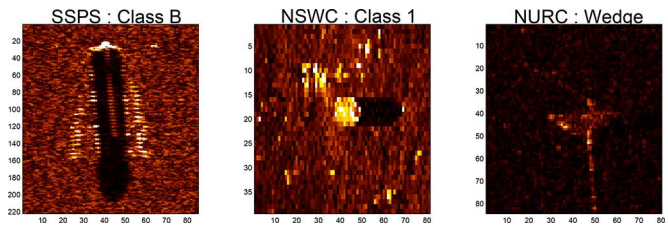


Fig. 3. Sample objects from three different databases: SSPS, NSWC, and NURC. The data illustrate the differing characteristics across objects and domains.

Additionally, we also suggest a robust segmentation technique based on MSC [19] for isolating the brightest and darkest regions in a sonar image, from the background. This technique is used for extracting objects and shadows from sonar images, followed by a denoising stage that exploits spatial coherence of the segmented pixels to prevent overclustering.

C. Outline

The block diagram in Fig. 4 shows the stages in reliability-aware classification. First, the images are segmented to detect the object and shadow. Then, Zernike moment magnitude features are computed on both shapes. Finally, a reliability-aware Bayesian network model is trained to discriminate between the object types. The pipeline is similar for the train and test images except for the training phase which uses class labels for the train images. Reliability of shadow features is estimated during training and is used to learn parameters of the Bayesian network model. Given shadow and object features on the test images, the model can now be used to output the class posteriors. Each of these stages is presented in detail in Sections III–VI.

III. SEGMENTATION

Although we restrict our interest to object classification in this paper, classification in underwater sonar images is typically preceded by object detection and segmentation stages. In this work, we assume that the position of object in the image is roughly known. Nevertheless, object segmentation is essential to locate the exact position of the object and extract information about

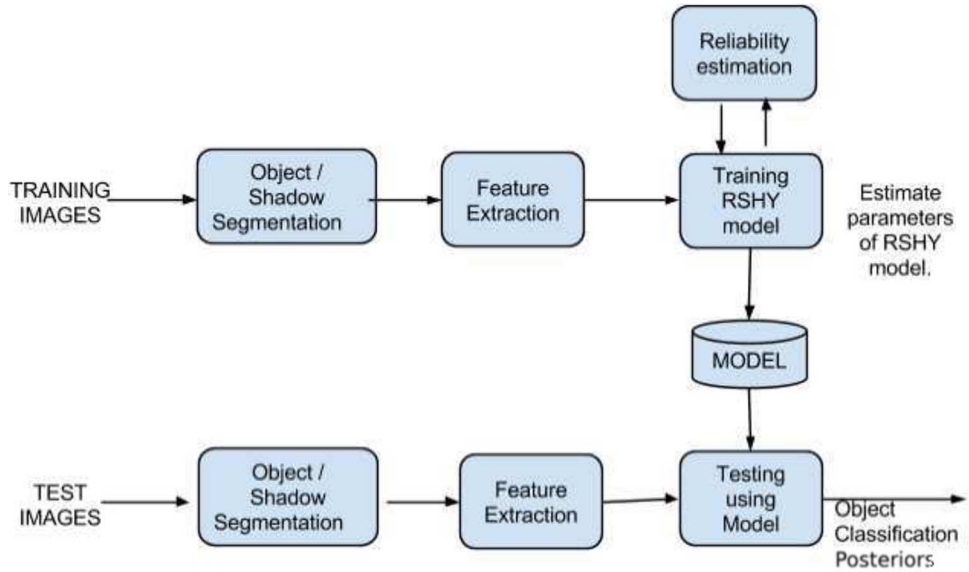


Fig. 4. Block diagram outlining steps in reliability-aware classification using shadow and object features.

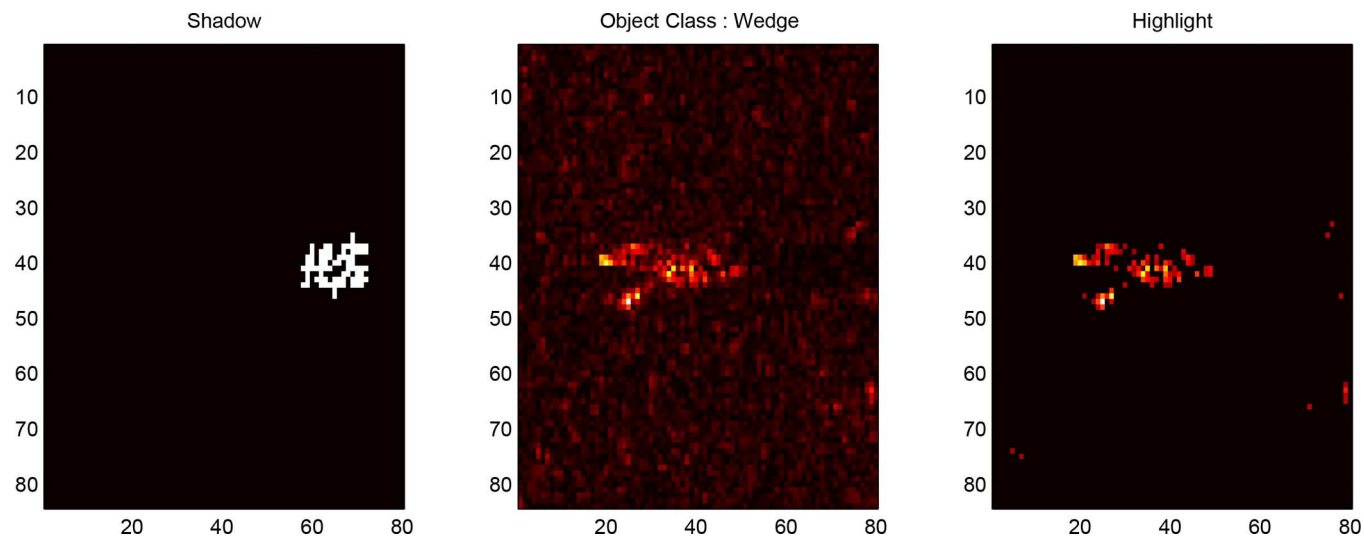


Fig. 5. Example from the NURC database showing highlight (right) and shadow (left) segmented from the original image (center) using the MSC-based segmentation scheme.

its shape. A similar segmentation technique is also required to extract auxiliary features from shadows.

Shadow segmentation in sonar images can be challenging at times owing to the nature of sonar imagery and other factors such as the texture of the seafloor or the beam angle, as can be seen in Fig. 5. These can either cause the shadow to be absent at times, or lead to incorrect segmentation resulting in unreliable shadow information on some of the images. When present, the shadows can sometimes be very discriminative, owing to their much larger size in comparison to the echo/object highlight. In other cases, a shadow can be either completely absent or lost among the clutter on the seafloor. This lack of reliable shadow information prevents fusion of shadows directly at the feature level and necessitates a scheme that can cope with the missing or noisy shadows for certain samples. This adaptive fusion scheme is discussed in further detail in Section VI.

On the NURC and SSPS databases, we employ a segmentation method based on MSC [19] for both objects and their shadows, followed by denoising. On the NSWC database, segmentation methods are only applied for extracting shadows, since the position of objects is precisely known. First, we describe the general segmentation algorithm followed by modifications to specify the algorithm to either segment highlights or shadows.

A. MSC-Based Segmentation

In this work, we adapt a segmentation scheme based on MSC [19]. MSC is used to facilitate segmentation by adaptively clustering the intensity values in the sonar images. MSC is a mode seeking algorithm that is popularly used in unsupervised clustering. Given initial points in a data distribution, the MSC algorithm ascends in the direction of the density gradient to the

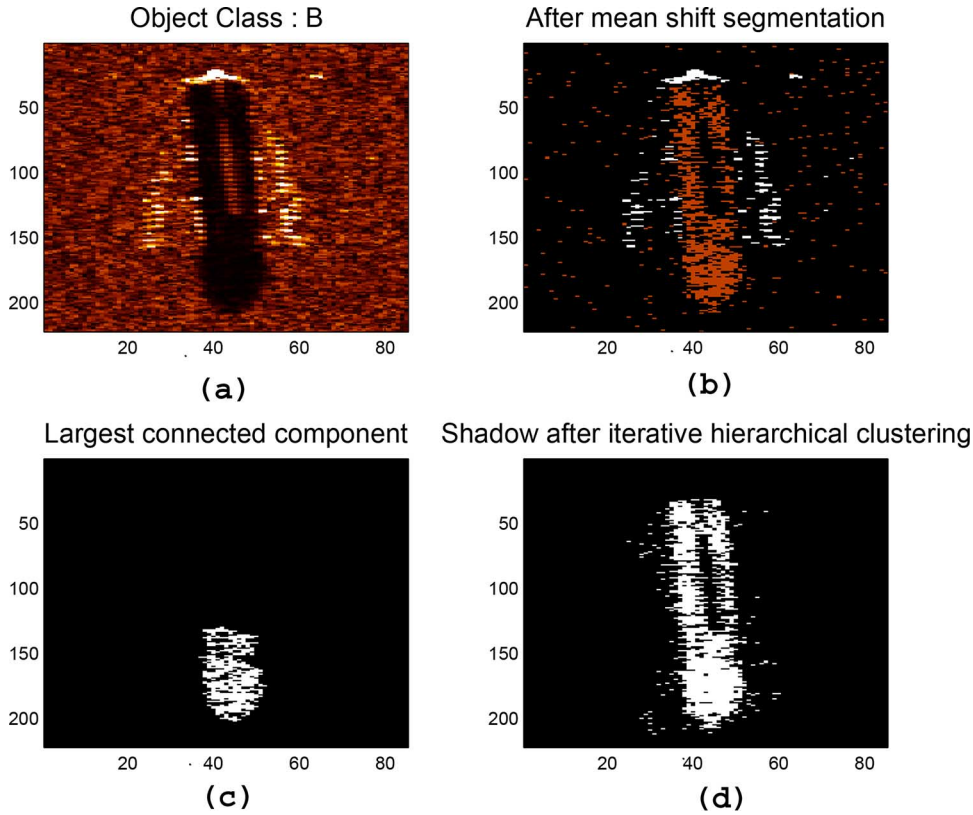


Fig. 6. Stages in segmentation of an image (a) from the SSPS database, (b) mean-shift segmentation into shadow (brown) and highlight (white), and (d) iterative hierarchical clustering (bottom-right). (c) Effect of doing the largest connected component analysis instead.

nearest local mode of the distribution. In general MSC, the density at a point can be measured using an arbitrary kernel function $K(x)$. For our purposes, in this paper, we choose a simple uniform kernel function, as shown in (1), that provides a good tradeoff between complexity and performance. We simply measure the density at a point in terms of the number of points within a circle of fixed radius R_b. This radius is referred to as the bandwidth in MSC.

The algorithm proceeds by computing the mean of all points within this specified radius (bandwidth) of the initial point v_0. This point is then moved to the mean in the next iteration, as shown in

$$v_{n+1} = \frac{\sum_p K(p - v_n)p}{\sum_p K(p - v_n)}, \qquad p \in \mathrm{dom}(v)$$
$$K(p - v) = \begin{cases} 1, & \text{if } \|p - v\| \leq R_b \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

This procedure is repeated until the mean converges to a local mode of the distribution. Convergence is achieved when the update in the mean is below a certain threshold, i.e., $\|v_{n+1} - v_n\| \leq \epsilon$. All points traversed in the process are then assigned to this mode/cluster.

This process is repeated for several random initial points until all the points in the image have been assigned to some cluster. The bandwidth parameter R_b provides control over how fine-grained the clustering is. A smaller bandwidth might lead to a large number of clusters that are similar to each other, while a larger bandwidth will merge the similar clusters. Since the clustering occurs in the colorspace, R_b has the same units as that of the color intensity.

B. Choosing the Segmentation Target: Highlight or Shadows

For our segmentation algorithm, we also find that changing the value of the bandwidth parameter R_b allows us to choose the object of segmentation to either the shadow or the highlight. Remember that a higher bandwidth ensures larger and more coherent clusters while a smaller bandwidth value yields smaller, more fragmented clusters. For example, to segment shadows, a smaller bandwidth is chosen, accompanied by the knowledge that the shadow regions have the lowest intensity values. We find that in all our 8-b image (0–255) data sets, a bandwidth $R_b = 10$ manages to differentiate the shadow from the background clutter. The bandwidth parameter can be easily tuned to the dynamic range of the images under consideration. If segmentation fails, it can be verified by setting a threshold on the ratio of the number of pixels in the segmented shadow to that in the image. An adaptive scheme is then used which uses a lower bandwidth $(R_b = 5)$ to obtain a better segmentation.

For segmenting the highlight, a larger bandwidth parameter $(R_b = 15)$ is used which fuses all regions of low intensity values, except the bright highlight into a single cluster, while the pixels belonging to the highlight are accumulated in all other clusters. This approach is based on the observation that the highlight in an image is typically much brighter compared to the shadow or clutter around the object. The algorithms are presented formally in Algorithm 1.

Algorithm 1: Shadow and highlight segmentation using MSC.

Input: Image $\mathbf{F}$ containing N pixels
Output: Shadow segmentation binary mask $\mathbf{F}^S$ of the same size as $\mathbf{F}$
Cluster $\mathbf{F}$ using MSC with $R_b = 10$. Suppose this yields d clusters.
Sort clusters in ascending order of mean intensity value: $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{d-1}$
if $\frac{|\mathcal{C}_0|}{N} < 0.5$ **then**
Shadow $\mathcal{S}' \leftarrow \mathcal{C}_0$
else
Cluster image using MSC with $R_b = 5$.
Sort clusters in ascending order of their mean intensity values : $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{d'-1}$
Shadow $\mathcal{S}' \leftarrow \mathcal{C}_0$
Encode the set of points in $\mathcal{S}'$ as a binary mask

$$\mathbf{F}^S \mid \mathbf{F}^S_{ij} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{S}' \\ 0, & \text{otherwise} \end{cases}$$

end if
Input: Image $\mathbf{F}$ containing N pixels
Output: Highlight segmentation binary mask $\mathbf{F}^H$ of the same size as $\mathbf{F}$
Cluster image $\mathbf{F}$ using MSC with $R_b = 15$. Suppose this yields q clusters.
Sort clusters in ascending order of their mean intensity values : $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{q-1}$
Highlight $\mathcal{H}' \leftarrow \mathcal{C}_1 \cup \mathcal{C}_2 \cup \ldots \mathcal{C}_{q-1}$
Encode the set of points in $\mathcal{H}'$ as a binary mask $\mathbf{F}^H$ as shown for $\mathbf{F}^S$ above.

$\mathcal{C}_i, \mathcal{S}',$ and $\mathcal{H}'$ used in the algorithm are sets of tuples of pixel indices corresponding to each cluster, the segmented shadow, and highlight, respectively. $\mathbf{F}^S$ and $\mathbf{F}^H$ are binary masks of the same size as the original image $\mathbf{F}$ encoding each segmentation.

C. Denoising Using Iterative Hierarchical Clustering

Although MSC manages to isolate the shadow and highlight to a large degree, the segmentation still contains some noise [Fig. 6(b)]. Noise occurs as a result of clustering in the color space, since a cluster is not required to be a contiguous mass spatially. Thus, the segmented shadow is often fragmented, which renders simple denoising approaches, like selecting the largest connected component, useless [Fig. 6(c)]. We perform denoising based on a hierarchical clustering method to deal with this problem.

Specifically, we use an agglomerative hierarchical clustering (AHC) [20] technique that initializes by assigning one cluster to each point. At each stage, coherent clusters are fused together creating a hierarchy that defines how each point is related to others. The clustering is stopped when a certain cutoff criterion on the consistency [21] of clusters is met. From one such round of clustering, we select the largest cluster and proceed to cluster it similarly. This iterative scheme is stopped when the current cluster cannot be divided into further clusters by AHC according to the consistency cutoff parameter provided. We use an empirically determined consistency cutoff value of 1.5 for our experiments. The algorithm is stated in Algorithm 2. $\mathcal{C}_i$'s are used to denote sets of tuples of pixel indices belonging to each cluster as before.

Algorithm 2: Iterative hierarchical clustering algorithm used for denoising the segmentation

Input: MSC segmentation $\mathcal{S}'$
Output: Denoised segmentation $\mathcal{S}_{\text{ahc}}$ and shadow binary mask $\mathbf{F}^S$
$\mathcal{S}_{\text{ahc}} \leftarrow \mathcal{S}'$
repeat
Cluster $\mathcal{S}_{\text{ahc}}$ using AHC. Suppose this yields q clusters.
Sort clusters in descending order by number of points as $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{q-1}$
$\mathcal{S}_{\text{ahc}} \leftarrow \mathcal{C}_0$
until $q > 1$
Encode the set of points in $\mathcal{S}_{\text{ahc}}$ as a binary mask

$$\mathbf{F}^S \mid \mathbf{F}^S_{ij} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{S}_{\text{ahc}} \\ 0, & \text{otherwise} \end{cases}$$

The segmentation obtained after the above denoising is used to extract features from corresponding parts of the image.

IV. Feature Extraction

The role of feature extraction in pattern recognition problems is critical. Given the object heterogeneity and data uncertainty common in real world applications, the general consensus in practice is that there is no universally best feature for a problem. The general approach hence has been to find features tuned to a particular domain, and it is no different for underwater sonar images [22].

For the purposes of object classification, we desire that the algorithm be insensitive to small variations in intensity and orientation of the image. Some methods try to mitigate such variations within the classification algorithms [23]–[25], while others deal with them at the feature level [26], [27]. In addition, since most classification algorithms suffer from the curse of dimensionality [28], we would also like to characterize the object as compactly as possible. Zernike moments meet the above needs and are popularly used as invariant features for object classification and shape-based content retrieval tasks [29]–[31].

A. Zernike Moments

Zernike moments of an image are computed via an orthogonal transform in the polar domain, with the degree of the representation controlling the degree of generalizability. We use magnitudes of Zernike moments, which have been shown to possess rotational invariance properties for object recognition [29]. Their robustness to variabilities in underwater images has also been well established [32], [33]. To compute Zernike moments ζ_{nm} of order (n, m), we find the projection of the image $\mathbf{F}$ with the basis function $\mathbf{V}_{nm}$ as shown in (4). To simplify notation,

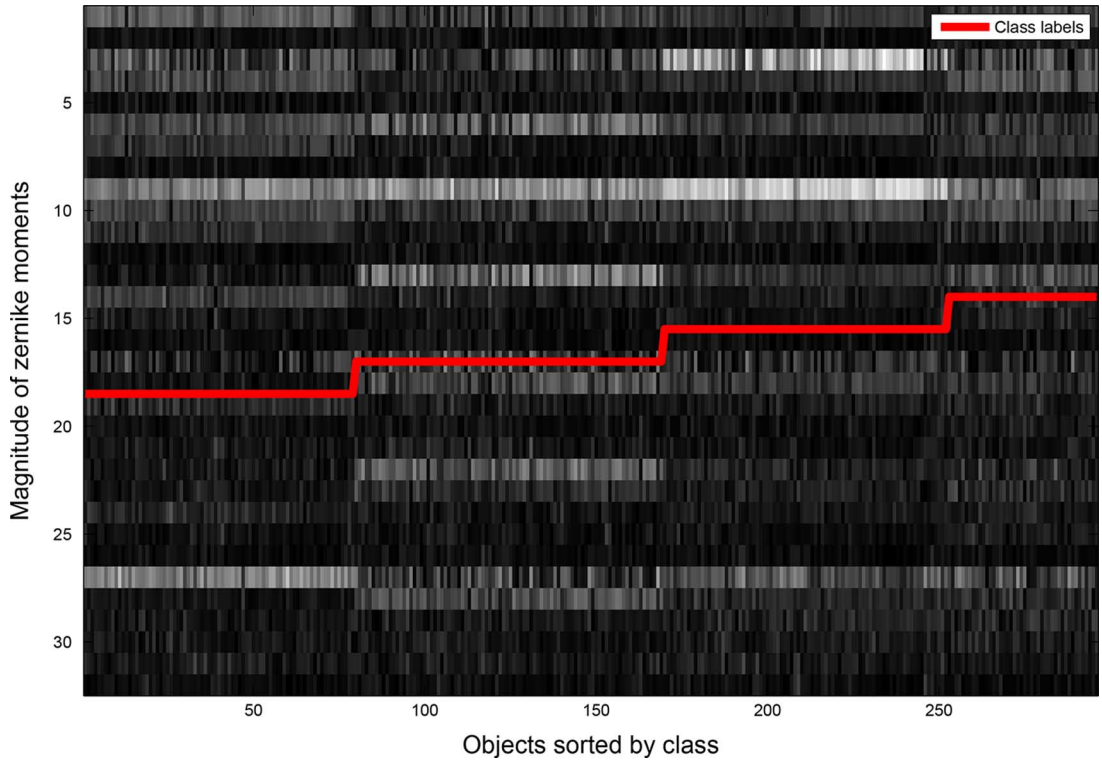


Fig. 7. Zernike moment magnitude features for 296 samples from the NSWC database. The solid red line is an indicator of the class labels and the vertical axis corresponds to different feature dimensions. Note how features are similar in one class in spite of the objects being oriented at different angles and varying across different classes.

we restrict ourselves to square images. In addition, we transform the coordinate system of the image such that the origin is at the center of mass and the entire object just fits within the unit circle $a^2+b^2=1$. a and b are related to their polar counterparts ρ and θ as $a=\rho\cos(\theta), b=\rho\sin(\theta), \rho\leq 1$.

$$\mathbf{V}_{nm}(a,b)=\mathbf{V}_{nm}(\rho,\theta)=\underline{w}_{nm}(\rho)e^{-jm\theta} \tag{2}$$

$$\underline{w}_{nm}(\rho)=\sum_{s=0}^{\frac{n-|m|}{2}}(-1)^s\frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}\rho^{n-2s} \tag{3}$$

$$\zeta_{nm}(\mathbf{F})=\frac{n+1}{\pi}\sum_x\sum_y\mathbf{F}(a,b)\mathbf{V}^*_{nm}(a,b),\quad a^2+b^2\leq 1 \tag{4}$$

where $0\leq n\leq N, |m|\leq n, n-|m|$ is even.

The range of n selects the order of Zernike moments and the degree of generalizability of the description. From our previous work [2] with sidescan sonar images, we found that representations of the order $N=10$ contain sufficient information for object classification that also generalizes well, yielding 36 unique moments satisfying the constraints above. Note that after rotation by an angle α, only the phases of the Zernike moments depend on the object's orientation. If the Zernike moments after rotation of the object by an angle α are denoted by ζ^{α}_{nm}, then the relation to the original set of Zernike moments ζ^{0}_{nm} is shown in

$$\zeta^{\alpha}_{nm}=\zeta^{0}_{nm}e^{-j\alpha}. \tag{5}$$

Thus, the magnitudes of Zernike moments are rotationally invariant.

This property can also be seen in Fig. 7 which shows the magnitude of Zernike moments corresponding to objects grouped together by their class. We extract Zernike moments from the pixels corresponding to the highlight [$\mathbf{F}\odot\mathbf{F}^H$ ($\odot$ indicates entrywise product)] and compute their magnitude. The thick red line on the graph indicates the true class label of the object. The similarity of Zernike magnitude features within a class is evident in spite of different objects being in different orientations. To formally verify this, we performed an analysis-of-variance (ANOVA) test to check the hypothesis that the variance of features within a class (Σ_{within}) is smaller than the variance of features between classes (Σ_{between}). The alternate hypothesis was $\text{tr}(\Sigma_{\text{within}})<\text{tr}(\Sigma_{\text{between}})$ and the test statistic $\mathcal{F}$ is shown in (9). Rows of the matrix $Q(k)$ contain features ζ only for samples from the class k.

$$\mu_k=\mathbb{E}[Q(k)] \tag{6}$$

$$\Sigma_{\text{within}}=\sum_{k=1}^{K}\Sigma_k,\qquad \Sigma_k=\mathbb{E}\left[Q(k)_0Q(k)_0^T\right] \tag{7}$$

$$\Sigma_{\text{between}}=\text{Cov}(\mu_1,\mu_2,\ldots,\mu_K) \tag{8}$$

$$\mathcal{F}=\frac{\text{tr}(\Sigma_{\text{within}})}{\text{tr}(\Sigma_{\text{between}})}. \tag{9}$$

A p-value of 0.07 was obtained for the F-test which indicates that there is sufficient evidence to reject the null hypothesis in ANOVA that the class means are equal. In other words, the variance of features within a class is significantly smaller than the

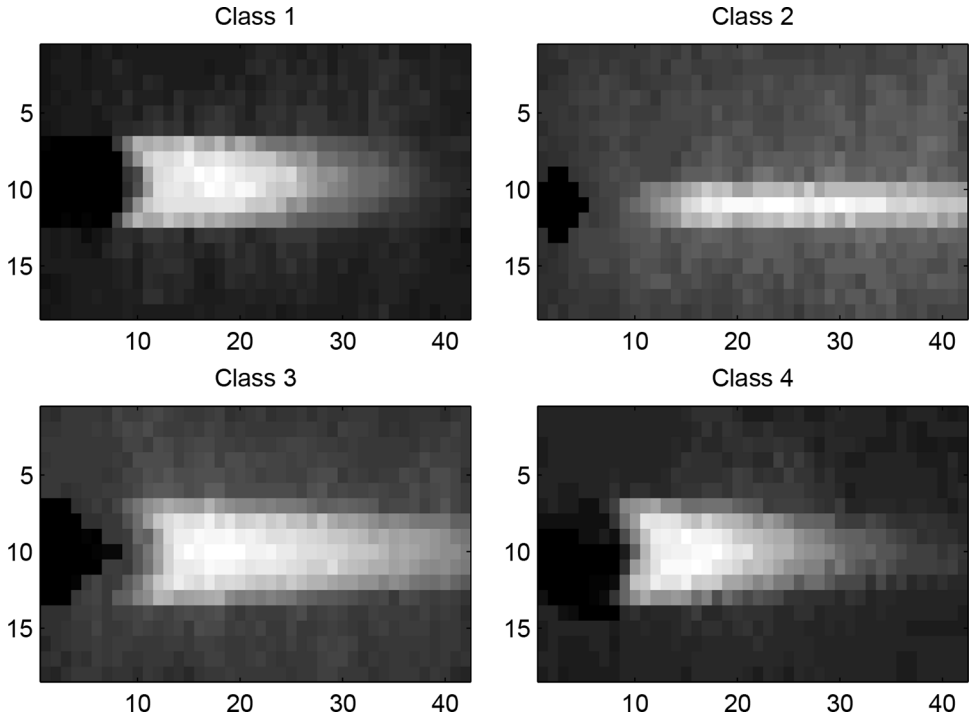


Fig. 8. Average shadows from each class in the NSWC database. (Negative of images shown to highlight shadows).

total variance across all classes. Hence, the feature should contain sufficient discriminative information.

V. Using Shadows for Object Classification

A. Shadow Features

While computing feature representations from shadows, intensity values are neglected. Only the shape represented by the binary mask $\mathbf{F}^S$ is encoded using the previously mentioned Zernike features. One of the major challenges in using shadow and object features in the same framework is that the features extracted from shadows are often not as reliable as those from objects. This can be either because shadow features are missing for some objects or, if they are present, do not provide sufficient discrimination for classification tasks by themselves. This can be seen in Fig. 8, for the average shadows extracted on the NSWC data set for each class. Note that it is harder to spot differences between the shadows for classes 1, 3, and 4, while class 2 easily stands out and might help improve classification. This observation suggests that the shadow features are useful only in certain cases to resolve the ambiguities posed by the object shapes, hence making reliability-aware fusion of feature sets important.

Hence, we propose a reliability-aware feature fusion scheme. First, we compute the features $\underline{\mathbf{s}} = \zeta(\mathbf{F}^S), \underline{\mathbf{h}} = \zeta(\mathbf{F} \odot \mathbf{F}^H)$ for all samples, but assume that some of the samples may not have equally reliable shadow and highlight features. This reliability is expressed in a data-driven fashion. We test this idea by comparing it against the naive fusion schemes described next.

B. Naive Feature Fusion Results

First, we present results of naive feature fusion using popular classifiers, viz., support vector machine (SVM) and logistic regression (Table II). These fusion schemes do not take into account the reliability of the shadow features and attempt to blindly fuse the two feature sets. The fusion methods are implemented via the WEKA toolkit [34]. In Section VII, we also compare against a score fusion scheme.

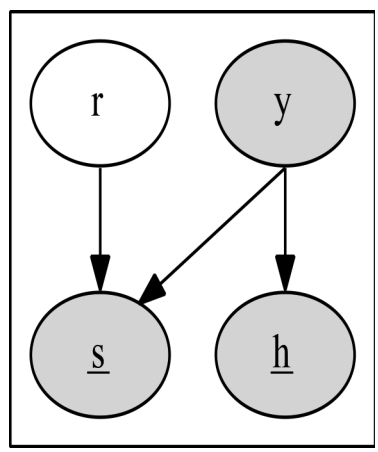


Fig. 9. Reliability node $\mathbf{r}$ acts as a switch for the link between class label $\mathbf{y}$ and the shadow features $\underline{\mathbf{s}}$. $\underline{\mathbf{h}}$ refers to the object highlight features. The shaded nodes represent a variable that can be directly observed whereas the unshaded node corresponding to $\mathbf{r}$ is assumed to be hidden. Also the arrows between pair of nodes indicate the only direct conditional dependencies.

C. Reliability-Aware Feature Fusion

Next, we provide a more concrete definition of reliability for classification. Suppose the reliability associated with each sample is known in advance. Then, the process of reliability-aware feature fusion consists of selectively using the reliable samples to train a reliable model, which is used to predict class labels. Finally, fusion must also be only performed for the reliable samples on the test set. Though this notion of reliability as selecting instances of a data set is easy

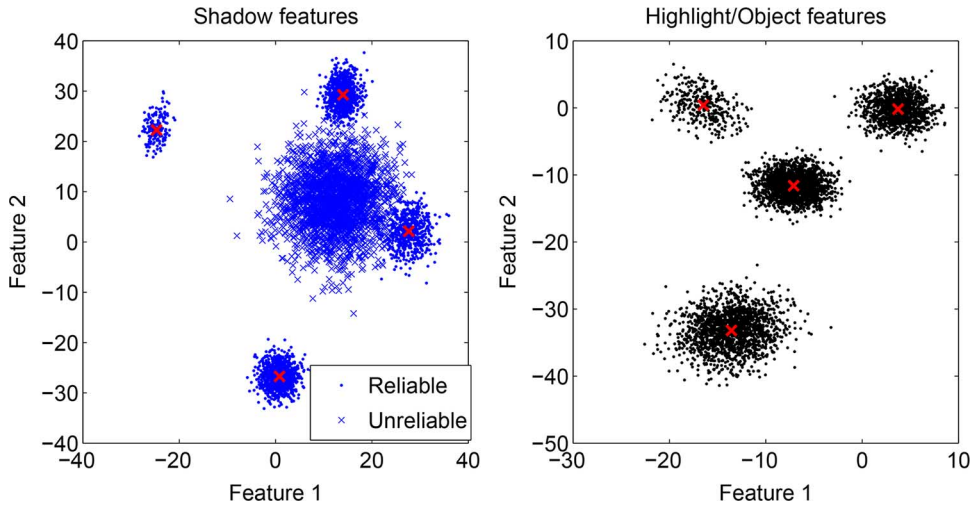


Fig. 10. Data generated using the model in Fig. 9. Red crosses indicate the centers for each of the class clusters. For generating the shadow features, $P(\mathbf{r}) = 0.5$ was used, i.e., 50% of the samples were randomly selected to have unreliable shadow features.

TABLE II
DIRECT FEATURE LEVEL FUSION DOES NOT CONSISTENTLY OFFER SIGNIFICANT IMPROVEMENT TO CLASSIFICATION ACCURACY BECAUSE OF VARIED RELIABILITY OF SHADOW FEATURES

Database	Features	Classifier	Accuracy(in %)
NSWC	Shadow	SVM	54.7
NSWC	Object	SVM	94.6
NSWC	Shadow+Object	SVM	95.3
NSWC	Shadow	Logistic	58.8
NSWC	Object	Logistic	93.6
NSWC	Shadow+Object	Logistic	93.6
NURC	Shadow	SVM	49.4
NURC	Object	SVM	57.5
NURC	Shadow+Object	SVM	61.9

to understand, it is not as clear what it means for the features to be reliable for a sample.

Since we are currently concerned with a classification task, we define reliability using the relation between features and class labels. In this work, we will make the assumption that when features are unreliable, they are generated by a random noisy model that does not depend on the class to which the object belongs. This is a common assumption made in many machine learning tasks such as automatic speech processing to deal with outliers [12], [13]. In Section VI, we present this assumption more formally using a Bayesian network.

VI. BAYESIAN NETWORK MODEL FOR RELIABILITY

To formulate the reliability-aware fusion problem as a Bayesian network we assume that there exists a latent binary random variable $\mathbf{r}$ that lets us choose how the partially reliable feature is generated. We assume the following structure for the graphical model, and make a naive Bayes assumption on the shadow ($\underline{\mathbf{s}}$) and highlight ($\underline{\mathbf{h}}$) features. In addition, we assume that the class-conditional probability for shadow features is a mixture model of reliable and unreliable features, as shown in (11).

Using local Markov properties, the joint distribution can be factorized as follows:

$$P(\mathbf{r}, \underline{\mathbf{s}}, \underline{\mathbf{h}}, \mathbf{y}) = P(\mathbf{y}) P(\underline{\mathbf{h}}|\mathbf{y}) P(\underline{\mathbf{s}}|\mathbf{r}, \mathbf{y}) P(\mathbf{r}) \tag{10}$$

where $\underline{\mathbf{s}}$ and $\underline{\mathbf{h}}$ are the shadow and highlight features, with the dimensions d_s and d_h, respectively; $\underline{\mathbf{s}} \in \mathbb{R}^{d_s}, \underline{\mathbf{h}} \in \mathbb{R}^{d_h}, \mathbf{y} \in \{1 \dots k\}, \mathbf{r} \in \{0, 1\}$. This model will be referred to as the RSHY model.

We use the reliability of features to refer to their dependence with respect to class labels, i.e., $P(\underline{\mathbf{s}}|\mathbf{y}, \mathbf{r} = 0) = P(\underline{\mathbf{s}})$. Thus, for $\mathbf{r} = 0$, i.e., when the shadow feature is completely unreliable, the model says that the shadow feature was generated independent of the class label, while $\mathbf{r} = 1$ indicates that the shadow feature is generated from a class-dependent mixture distribution or $P(\underline{\mathbf{s}}|\mathbf{y}, \mathbf{r} = 1) = P(\underline{\mathbf{s}}|\mathbf{y})$. Another way to think about this model is that it smoothly interpolates between the conditional dependent and independent models for shadow features

$$P(\underline{\mathbf{s}}|\mathbf{r}, \mathbf{y}) = P(\underline{\mathbf{s}}|\mathbf{y})^{\mathbf{r}} P(\underline{\mathbf{s}})^{1-\mathbf{r}}. \tag{11}$$

For simplicity, we assume all marginal and conditional distributions in this model to be Gaussian, which gives rise to three sets of Gaussian parameters. The generative process for each object can be described as shown in Algorithm 3.

Algorithm 3: Generative process for the RSHY model

Choose class label $\mathbf{y} \sim$ *Categorical*$(\eta_1, \eta_2, \dots, \eta_K)$
If $\mathbf{y} = k$, choose highlight features $\underline{\mathbf{h}} \sim$ *Gaussian*$(\mu_k^{\Theta}, \Sigma_k^{\Theta})$
Choose $\mathbf{r} \sim$ *Bernoulli*(ρ)
if $\mathbf{r} = 0$ **then**
  Choose shadow features $\underline{\mathbf{s}} \sim$ *Gaussian*$(\mu^{\Phi}, \Sigma^{\Phi})$
else
  Choose shadow features $\underline{\mathbf{s}} \sim$ *Gaussian*$(\mu_k^{\Omega}, \Sigma_k^{\Omega})$ if $\mathbf{y} = k$
end if

The model parameters are defined as follows:

- η_k: prior probabilities for each class;
- ρ: probability of shadow features for a sample being reliable on average;
- Θ_k: Gaussian model for highlight feature class-conditional probability;
- Ω_k: Gaussian model for shadow feature class-conditional probability;
- Φ: Gaussian model for shadow feature marginal probability.

A. Samples Generated by the Model

Before diving into further details it might be useful to demonstrate, via an example, the data characteristic that the model assumes. This is particularly simple in our case, since our model is a generative model, and can be easily sampled from, such as using random parameter settings as adopted here. Fig. 10 shows a distribution of $N = 1000$ feature points generated using the model. We restrict the dimensionality to 2 for ease of interpretation. Note that the set of unreliable samples for shadow features are expected to have larger scatters in the feature space and their distribution does not depend on what class they belong to. This figure also illustrates the possible issues in parameter estimation in case the patterns from two classes or the "garbage model" have a significant overlap.

B. Maximum Likelihood Parameter Estimation

We estimate the parameters of the proposed Bayesian network model using the EM algorithm [35]. The variable $\mathbf{r}$ is a hidden parameter in this model. For ease of analysis, we will represent the class labels $\mathbf{y}$ using the 1-of-K encoding, i.e., if $\mathbf{y}_i = k$ originally, we use the notation $\sum_k \mathbf{y}_{ik} = 1, \mathbf{y}_{ik} = 1$ to indicate that the ith sample belongs to the kth class. Ideally, we would like to maximize the following log-likelihood:

$$\begin{aligned}\mathcal{L} &= \log\left(\prod_{i=1}^{N} P\left(\mathbf{r}_i, \underline{\mathbf{s}}_i, \underline{\mathbf{h}}_i, \mathbf{y}_i\right)\right) \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\left(\log\eta_k + r_i \log\mathcal{N}(s_i;\Omega_k) + \log\mathcal{N}(h_i;\Theta_k)\right) \\ &\quad + r_i\log\rho + (1-r_i)\left(\log\mathcal{N}(s_i;\Phi) + \log(1-\rho)\right) \qquad (12)\end{aligned}$$

where $\mathcal{N}(X, \Psi)$ refers to the probability according to the multivariate normal distribution defined by the model Ψ.

1) E-Step: Since r_i for each sample is a hidden random variable, we compute the following posterior distribution f:

$$P\left(\mathbf{r}_i|\underline{\mathbf{s}}_i, \underline{\mathbf{h}}_i, \mathbf{y}_i\right) = \frac{P\left(\mathbf{r}_i, \underline{\mathbf{s}}_i, \underline{\mathbf{h}}_i, \mathbf{y}_i\right)}{\sum_{\mathbf{r}_i=0,1} P\left(\mathbf{r}_i, \underline{\mathbf{s}}_i, \underline{\mathbf{h}}_i, \mathbf{y}_i\right)}. \qquad (13)$$

This is used to compute the expected value of the log-likelihood function. First, we compute a soft value for each r_i, which represents the uncertainty in our knowledge of the hidden variable

$$\begin{aligned}\gamma_i &= \mathbb{E}_f(\mathbf{r}_i) = P\left(\mathbf{r}_i = 1|\underline{\mathbf{s}}_i, \underline{\mathbf{h}}_i, \mathbf{y}_i\right) \\ &= \frac{\rho\prod_{k=1}^{K}\mathcal{N}(s_i;\omega_k)^{Y_{ik}}}{\rho\prod_{k=1}^{K}\mathcal{N}(s_i;\omega_k)^{Y_{ik}} + (1-\rho)\mathcal{N}(s_i;\Phi)}. \qquad (14)\end{aligned}$$

Then, the expected log-likelihood on f is given as

$$\begin{aligned}\mathcal{L}' &= \sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\left(\log\eta_k + \gamma_i\log\mathcal{N}(s_i;\Omega_k) + \log\mathcal{N}(h_i;\Theta_k)\right) \\ &\quad + \gamma_i\log\rho + (1-\gamma_i)\left(\log\mathcal{N}(s_i;\Phi) + \log(1-\rho)\right). \qquad (15)\end{aligned}$$

2) M-Step: Now since the hidden variable r_i has been replaced by its expected value γ_i, we can directly calculate the

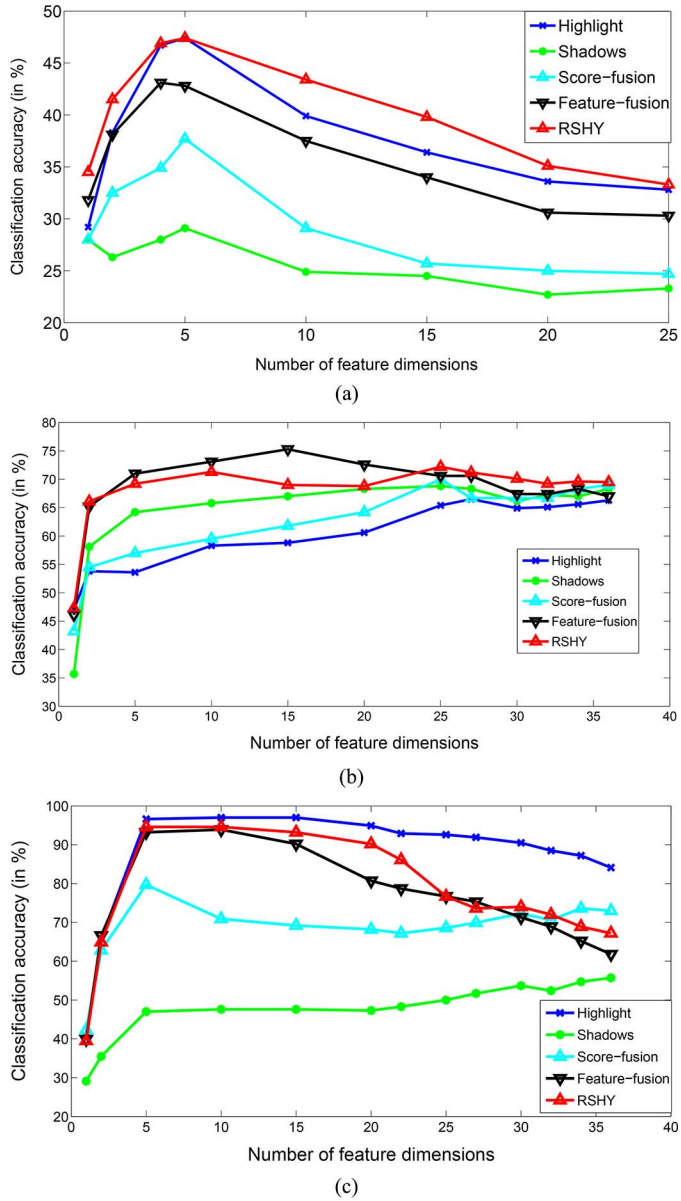


Fig. 11. Comparison of classification accuracies on three different databases using different feature sets and fusion schemes. Reliability-aware fusion yielded the best performance on the NURC data set where the shadow features were most unreliable (see $\hat{\rho}$ values). Also, in general dimension reduction helps since parameters of the Gaussian models can be better estimated. (a) NURC database ($\hat{\rho} \approx 0.3$). (b) SSPS database ($\hat{\rho} \approx 0.7$). (c) NSWC database ($\hat{\rho} \approx 0.4$).

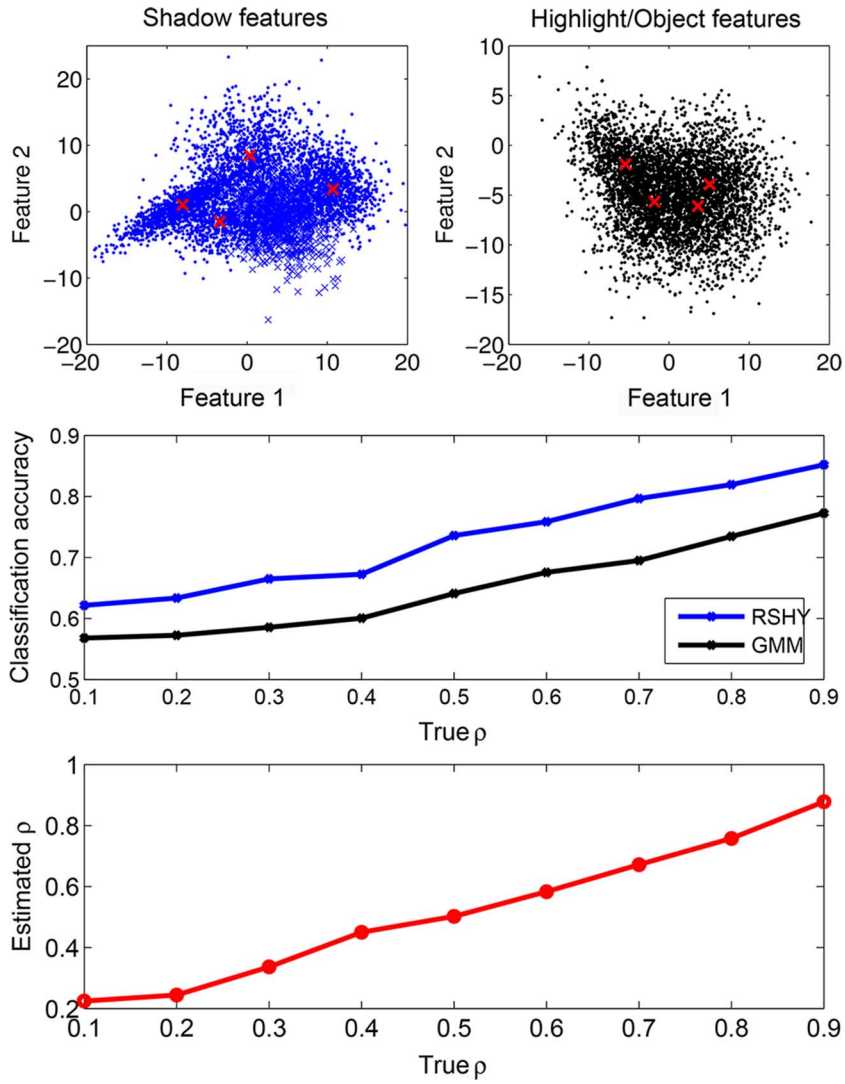


Fig. 12. Performance degradation with change in reliability ρ. Accuracies indicated are on the 5000 samples synthetically generated data set. The features presented are for $\rho = 0.8$. GMM indicates training a Gaussian model per class and classification using a MAP rule.

ML estimates of the above parameters. The constrained optimization is changed to an unconstrained optimization problem using the method of Lagrange multipliers and the following intuitive update equations are obtained for each of the parameters:

$$\eta_k = \frac{\sum_{i=1}^{N} y_{ik}}{N} \tag{16}$$

$$\rho = \frac{\sum_{i=1}^{N} \gamma_i}{N} \tag{17}$$

$$\mu_k^{\Theta} = \frac{\sum_{i=1}^{N} y_{ik} h_i}{\sum_{i=1}^{N} y_{ik}}$$

$$\Sigma_k^{\Theta} = \frac{\sum_{i=1}^{N} y_{ik}(h_i - \mu_k^{\Theta})(h_i - \mu_k^{\Theta})^T}{\sum_{i=1}^{N} y_{ik}} \tag{18}$$

$$\mu_k^{\Omega} = \frac{\sum_{i=1}^{N} \gamma_i y_{ik} s_i}{\sum_{i=1}^{N} \gamma_i y_{ik}}$$

$$\Sigma_k^{\Omega} = \frac{\sum_{i=1}^{N} \gamma_i y_{ik}(s_i - \mu_k^{\Omega})(s_i - \mu_k^{\Omega})^T}{\sum_{i=1}^{N} \gamma_i y_{ik}} \tag{19}$$

$$\mu^{\Phi} = \frac{\sum_{i=1}^{N}(1-\gamma_i) s_i}{\sum_{i=1}^{N}(1-\gamma_i)}$$

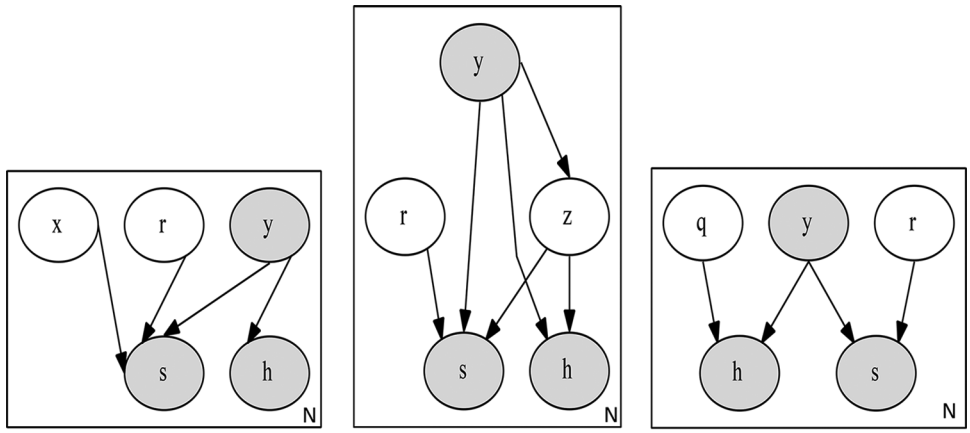
$$\Sigma^{\Phi} = \frac{\sum_{i=1}^{N}(1-\gamma_i)(s_i - \mu^{\Phi})(s_i - \mu^{\Phi})^T}{\sum_{i=1}^{N}(1-\gamma_i)}. \tag{20}$$

Note that the weighting terms for the update equations for models Θ, Ω, and Φ provide some intuition as to what they represent. Thus, Θ parameters are learned on highlight features per class, while Ω or Φ is learned selectively on shadow features from the reliable or unreliable samples. We iterate the E- and M-steps back and forth until the data log-likelihood converges. The EM algorithm ensures that the data log-likelihood is non-decreasing after each iteration.

Note that only parameters $\rho, \mu_k^{\Omega}, \Sigma_k^{\Omega}, \mu^{\Phi}$, and Σ^{Φ} need to be updated at each iteration, while parameters η_k, μ_k^{Θ}, and Σ_k^{Θ} are only estimated once.

C. Inference of Posterior Distributions of Unknown Variables

Once the parameters are estimated, we can infer the posterior distribution for class labels or reliability for the test samples.



(a) (b) (c)

Fig. 13. Suggested modifications to the RSHY model by using a GMM for each class and the garbage model. QRSHY adds unreliability to both feature modalities. (a) RXSHY: GMM for garbage model Φ only. (b) RZSHY: GMM models Θ, Ω, Φ. (c) QRSHY: Doubly unreliable model.

1) Inferring Class Labels: For classification, we can now directly estimate the posterior values for each class given the shadow and highlight features

$$P(\mathbf{y}_{ik}=1|\underline{\mathbf{s}}_i,\underline{\mathbf{h}}_i) = \frac{\sum_{\mathbf{r}=0,1} P(\mathbf{r},\underline{\mathbf{s}}_i,\underline{\mathbf{h}}_i,\mathbf{y}_{ik}=1)}{\sum_{j=1}^{K}\sum_{\mathbf{r}=0,1} P(\mathbf{r},\underline{\mathbf{s}}_i,\underline{\mathbf{h}}_i,\mathbf{y}_{ij}=1)} \tag{21}$$

$$= \frac{\eta_k \mathcal{N}(h_i;\Theta_k)\left(\rho\mathcal{N}(s_i;\Omega_k)+(1-\rho)\mathcal{N}(s_i;\Phi)\right)}{\sum_{j=1}^{K}\eta_j \mathcal{N}(h_i;\Theta_j)\left(\rho\mathcal{N}(s_i;\Omega_j)+(1-\rho)\mathcal{N}(s_i;\Phi)\right)}. \tag{22}$$

2) Estimating Reliability of a Sample: Alternatively, it is also possible to infer the reliability of a given sample's features as follows:

$$P(\mathbf{r}_i|\underline{\mathbf{s}}_i,\underline{\mathbf{h}}_i) = \frac{\sum_{k=1}^{K} P(\mathbf{r}_i,\underline{\mathbf{s}}_i,\underline{\mathbf{h}}_i,\mathbf{y}_{ik}=1)}{\sum_{\mathbf{r}_i=0,1}\sum_{k=1}^{K} P(\mathbf{r}_i,\underline{\mathbf{s}}_i,\underline{\mathbf{h}}_i,\mathbf{y}_{ik}=1)} \tag{23}$$

$$P(\mathbf{r}_i=1|\underline{\mathbf{s}}_i,\underline{\mathbf{h}}_i) = \frac{\sum_{k=1}^{K}\eta_k \mathcal{N}(h_i;\Theta_k)\rho\mathcal{N}(s_i;\Omega_k)}{\sum_{k=1}^{K}\eta_k \mathcal{N}(h_i;\Theta_k)\left(\rho\mathcal{N}(s_i;\Omega_k)+(1-\rho)\mathcal{N}(s_i;\Phi)\right)} \tag{24}$$

While the inference of class labels in this Bayesian network framework does not require the reliability posterior distribution for samples on the test set to be estimated explicitly, they could be used with another classifier to weight samples during training. The reliabilities for samples in the training set can be simply obtained by the estimated γ_i values at the end of all iterations. Also, note that although the reliability variable $\mathbf{r}$ was only tied to the shadow features, inference involves both the shadow and highlight features. This is due to the "upstream" nature of inferencing.

VII. Experiments and Results

To test the efficacy of the above model, we use it for object classification on the sonar image data sets mentioned earlier. Feature level and score level fusion schemes, being among the most commonly used blind fusion strategies, are adopted as our baseline. The proposed Bayesian network model and the baseline Gaussian models are trained on the Zernike moment magnitude features described earlier. For a fair comparison, we choose identical baseline models, i.e., class-conditional Gaussian models (per class feature models). We train separate Gaussian models on shadow (Λ_s) and highlight (Λ_h) features and a joint model (Λ_{sh}) on both. Finally, we also compare against a uniformly weighted score level fusion which averages the posteriors obtained using the highlight and shadow models, as shown in Table III. Each Gaussian model is learned in a naive Bayes fashion, that is, the covariance matrices are assumed to be diagonal. This alleviates the data-sparsity problem to some extent. In addition, we perform dimension reduction using principal component analysis (PCA). For each of the five folds, the PCA transformation is learned on the training folds' data and used to transform data in the test fold. This reduces the number of parameters to be trained, improving the accuracy of the model. Results are presented for a different number of feature dimensions in Fig. 11.

TABLE III
Decision Rules for Different Classification Schemes for Which Results Are Presented in Fig. 11

Scheme	Predicted class posteriors
Shadows	$P(\mathbf{y}_{ik}=1\|\underline{\mathbf{s}}_i,\Lambda_s)$
Highlight	$P(\mathbf{y}_{ik}=1\|\underline{\mathbf{h}}_i,\Lambda_h)$
Feature Fusion	$P(\mathbf{y}_{ik}=1\|[\underline{\mathbf{s}}_i\ \underline{\mathbf{h}}_i],\Lambda_{sh})$
Score Fusion	$(P(\mathbf{y}_{ik}=1\|\underline{\mathbf{s}}_i,\Lambda_s)+P(\mathbf{y}_{ik}=1\|\underline{\mathbf{h}}_i,\Lambda_h))/2$
Rel. Aware Fusion	$P(\mathbf{y}_{ik}=1\|\underline{\mathbf{s}}_i,\underline{\mathbf{h}}_i,\Lambda_{rsh})$

All experiments are performed using fivefold cross validation in which the data set is divided into five parts/folds. The classification experiment is then performed five times, each time holding out one of the folds as test data, and training on the remaining four. The average classification accuracy is reported at the end, which ensures that no bias is introduced due to the train set. For the proposed Bayesian network model, once the graphical model (Λ_{rsh}) is trained, we infer posteriors for class labels, as described in Section VI-C1. We again make naive Bayes independence assumptions like earlier. Additionally, all features are z-score normalized so that each feature dimension

is zero-mean and unit variance. This ensures that any mismatch in magnitude between the training and test sets is minimized.

We compare the performance of the proposed algorithm on the NURC, SSPS, and NSWC data sets. From the results, we note that the proposed algorithm is useful to different degrees on different data sets (Fig. 11), due to the variable reliability of the shadow/object features in different data sets, as described in Section II. On the NURC data set, the algorithm performs better than traditional fusions schemes, owing to the inherent lack of reliability in the shadow features, which can be seen by the ρ value of approximately 0.3 (Table IV). In the SSPS data set, the shadow features are more reliable on average ($\rho \approx 0.7$), leading to a performance that is at par with the feature fusion results. However, in the NSWC data set, the highlight features are highly discriminative, owing to the precisely known object locations. Thus, all fusion approaches perform worse in comparison to using just the highlight features, on the NSWC data set. In general, the classification performance on a data set seems to be correlated with the reliability of features on that data set. When the shadow features on a data set are unreliable, the algorithm manages to improve over the baseline method of blind fusion.

While these results provide some intuition as to when reliability-aware classification might be useful, we would like to further explore the dependence of the model on this notion of feature reliability. We try to explain this in more detail in Section VIII.

VIII. Discussion

A. Synthetic Experiments: Accuracy Versus Reliability ρ

To understand the role of feature set reliability in fusion performance, we performed a simulation experiment by varying the degree of reliability ρ of the shadow features. We synthetically generated 2-D highlight and shadow features for $N = 5000$ samples using random parameters values in the graphical model as described earlier. Classification experiments were performed using a 60-40 split of the synthetic data using both the proposed and traditional schemes. By synthetically generating these features, we are able to control the average reliability ρ to study the classification robustness with change in data reliability. Compared to the example in Fig. 10 used for illustration, the data set generated for this experiment was made more realistic by adding significant overlap between the clusters from each class (Fig. 12). The separability can be adjusted during generation by controlling the ratio of intraclass to interclass variance (i.e., Fisher's criterion), and causes the accuracy to taper off for lower ρ when the data set is not easily separable. It is worth mentioning here that since the absolute value of accuracy is also contingent on the inherent separability of the classes (over which we often have little control in a real data set) it might be more useful to study the relative trends in the accuracy curve with change in ρ, rather than its absolute value.

The decrease in accuracy in this case (Fig. 12) results from a poor estimation of the garbage model. If the majority of the samples are unreliable, then a slight overlap of the reliable clusters in the data can cause the data samples to appear noise-like, causing the algorithm to incorrectly assign low weight to them during training. For the synthetic data experiment, we observe that the proposed method performs better than the reliability blind case (GMM).

TABLE IV
Average Reliability Estimates for Shadow Features According to Different Models. (Note the Relation Between the Model Complexity and ρ)

Model	$\hat{\rho}$	
RSHY	0.39	
RZSHY	0.60	
RXSHY	0.25	
QRSHY	shad	high
NURC	0.39	0.70
SPSS	0.68	0.55
NSWC	0.42	0.95

B. Reliability and Model Complexity

Although the Bayesian network models our reliability assumptions well, it suffers from the drawback that each class-conditional distribution is modeled as a multivariate Gaussian distribution which may not be appropriate for real data sets. As we discussed in Section VII, data sparsity additionally can force us to make strong assumptions such as diagonality of the covariance matrix. To address this issue, we substitute the models for each of the class-conditional distribution and the shadow random model Φ with more complicated models. Specifically, we model each of these distributions as GMMs. In the machine learning literature, GMMs are well known for their ability to model arbitrary multimodal distributions.

To test the role of model complexity in the estimation of reliability we perform the following modifications to the RSHY model. First, we replace the Gaussian model Φ used as the shadow random model, by a GMM $\{\Phi_j\}_{j=1}^{J}$ [Fig. 13(a)]. Second, we also change the reliable shadow model $\{\Omega_k\}_{k=1}^{K}$ and the highlight models $\{\Theta_k\}_{k=1}^{K}$ by a mixture of GMMs Ω_j^k and Θ_j^k.[1] A latent variable $\mathbf{z}$ now selects the mixture component inside each class-conditional distribution [Fig. 13(b)]. These modifications[2] also allow us to better understand the implications of the model assumptions. After training these models on the same data sets as before, we observe that in each case the estimated average reliability ρ depends on the complexity of the model (Table IV). A stronger model that is able to explain and fit the data better is assigned more weight causing the reliability to bend in its direction. In other words, these results suggest that it might be misleading to use these reliability values in isolation from the model they were learned on. This warrants a note of caution regarding interpreting, out of the context, the ρ values obtained by the proposed algorithm. Finally, we also associate a reliability latent variable Q with the highlight features [Fig. 13(c)], which lets us compare the reliability of the two feature sets simultaneously.

[1] Θ_j^k indicates the parameters for the jth mixture component for the kth class.

[2] Details of ML parameter estimation and inference for these models, although omitted from this paper, are straightforward to derive by adding another hidden variable $\mathbf{z}/\mathbf{x}$ in the E-step.

IX. Conclusion and Future Work

Owing to the varied reliability of shadow features, their direct fusion, as is typically done in pattern recognition, does not prove to be a robust approach in our case. Instead, for the weak feature set, each sample is weighted according to its worth for the task at hand. We proposed a reliability-aware Bayesian network graphical model to exploit this notion of reliability of a sample in terms of its dependence in distribution. Missing features are assumed to be generated at random from a garbage model. This approach allows us to combine the feature sets efficiently and extract additional classification performance from the otherwise unreliable shadow features. The proposed framework clearly establishes the case for reliability-aware classification of noisy feature modalities. In addition, we proposed a robust shadow/highlight segmentation technique using MSC. We also suggest a simple technique by exploiting the characteristics of a sidescan sonar image that allows us to choose the segmentation target by controlling the bandwidth parameter R_b.

Feature reliability r for a sample is modeled as an independent binary random variable depending on the parameter ρ, which is learned as an average reliability parameter over the training data set. This parametrization of feature reliability helps adapt to a given data set/domain while still allowing us to model reliability r as a function of the feature set. However, feature reliability might also depend on other factors such as the target variable or class label in this case. In the future, we would like to incorporate these dependencies in our model. We show the effect of varying model complexity in this framework, and conclude that the inferred degree of reliability is related to the complexity of the models used for shadow and highlight features.

The main shortcomings of the proposed model are that, being a generative model, it estimates more parameters than required for only the classification task. As a result, our model requires a large amount of data to reliably estimate the model parameters. In addition, the proposed reliability weighting in our method trains feature and background models on samples depending on their reliability. This selective use of samples can create further data sparsity, leading to issues such as singular covariance matrices during parameter estimation when the feature average reliability ρ is either very low or very high. Finally, in our current setup, the proposed method only shows benefit over traditional fusion on data sets where one of the feature has lower reliability.

Apart from providing a normalized quantitative measure for comparing the reliability of two feature sets [QRSHY: Fig. 13(c)], the only simple assumption in this framework [see (11)] can be easily extended to other models. Thus, in the future, we would like to extend our model to a discriminative framework with similar reliability assumptions, enabling us to use the limited data more efficiently. In this work, we assume average reliability to be a fixed parameter for training data. In the future, we would like to consider other factors in the model on which reliability might depend. Finally, we would like to test the efficacy of our model on other larger data sets.

References

[1] S. Reed, Y. Petillot, and J. Bell, "Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information," *IEE Proc.—Radar Sonar Navig.*, vol. 151, no. 1, pp. 48–56, 2004.

[2] N. Kumar, Q. Tan, and S. Narayanan, "Object classification in sidescan sonar images with sparse representation techniques," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2012, pp. 1333–1336.

[3] R. Garcia, T. Nicosevici, and X. Cufí, "On the way to solve lighting problems in underwater imaging," in *Proc. MTS/IEEE OCEANS Conf.*, vol. 2, pp. 1018–1024.

[4] J. S. Jaffe, K. D. Moore, J. McLean, and M. Strand, "Underwater optical imaging: Status and prospects," *Oceanography*, vol. 14, no. 3, pp. 66–76, 2001.

[5] G. J. Dobeck *et al.*, "Automated detection and classification of sea mines in sonar imagery," in *Proc. AeroSense*, 1997, pp. 90–110.

[6] J. C. Hyland and G. J. Dobeck, "Sea mine detection and classification using side-looking sonar," in *Proc. SPIE Symp. OE/Aerosp. Sens. Dual Use Photon.*, 1995, pp. 442–453.

[7] T. Aridgides, M. F. Fernandez, and G. J. Dobeck, "Side-scan sonar imagery fusion for sea mine detection and classification in very shallow water," in *Proc. Aerosp./Defense Sens. Simul. Controls*, 2001, pp. 1123–1134.

[8] T. Aridgides, D. Antoni, M. F. Fernandez, and G. J. Dobeck, "Adaptive filter for mine detection and classification in side-scan sonar imagery," in *Proc. SPIE Symp. OE/Aerosp. Sens. Dual Use Photon.*, 1995, pp. 475–486.

[9] S. Reed, Y. Petillot, and J. Bell, "Unsupervised mine detection and analysis in side-scan sonar: A comparison of Markov random fields and statistical snakes," in *Proc. Comput.-Aided Detection/Comput.-Aided Classification Conf.*, 2001.

[10] E. Dura, Y. Zhang, X. Liao, G. Dobeck, and L. Carin, "Active learning for detection of mine-like objects in side-scan sonar imagery," *IEEE J. Ocean. Eng.*, vol. 30, no. 2, pp. 360–371, Apr. 2005.

[11] G. Dobeck, "Algorithm fusion for automated sea mine detection and classification," in *Proc. MTS/IEEE OCEANS Conf. Exhibit.*, 2001, vol. 1, pp. 130–134.

[12] J. G. Wilpon, L. R. Rabiner, C.-H. Lee, and E. Goldman, "Automatic recognition of keywords in unconstrained speech using hidden Markov models," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 11, pp. 1870–1878, Nov. 1990.

[13] J. Caminero, D. De La Torre, L. Villarrubia, C. Martin, and L. Hernández, "On-line garbage modeling with discriminant analysis for utterance verification," in *Proc. 4th Int. Conf. Spoken Lang.*, 1996, vol. 4, pp. 2111–2114.

[14] J. Liu and X. Zhu, "Utterance verification based on dynamic garbage evaluation approach," in *Proc. IEEE 5th Int. Conf. Signal Process.*, 2000, vol. 2, pp. 819–822.

[15] H. Bourlard, B. D'hoore, and J.-M. Boite, "Optimizing recognition and rejection performance in wordspotting systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 1994, vol. 1, pp. 373–376.

[16] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, vol. 5, no. 2, pp. 197–227, 1990.

[17] J. A. Fawcett *et al.*, "Multi-aspect computer-aided classification of the citadel trial side-scan sonar images," *DRDC Atlantic Technical Memorandum*, vol. 29, 2008.

[18] V. Myers and J. Fawcett, "A template matching procedure for automatic target recognition in synthetic aperture sonar imagery," *IEEE Signal Process. Lett.*, vol. 17, no. 7, pp. 683–686, Jul. 2010.

[19] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.

[20] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY, USA: Wiley, 2012, ch. 10.9.2.

[21] C. T. Zahn, "Graph-theoretical methods for detecting and describing gestalt clusters," *IEEE Trans. Comput.*, vol. C-20, no. 1, pp. 68–86, Jan. 1971.

[22] J. Stack, "Automation for underwater mine recognition: Current trends and future strategy," in *Proc. SPIE*, 2011, vol. 8017, DOI: 10.1117/12.884475.

[23] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[24] J. Revaud, G. Lavoué, and A. Baskurt, "Improving Zernike moments comparison for optimal similarity and rotation angle retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 627–636, Apr. 2009.

[25] R. P. Wurtz, "Object recognition robust under translations, deformations, and changes in background," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 769–775, Jul. 1997.
[26] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.
[27] L. A. Torres-Mendez, J. C. Ruiz-Suarez, L. E. Sucar, and G. Gomez, "Translation, rotation, and scale-invariant object recognition," *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, vol. 30, no. 1, pp. 125–130, Feb. 2000.
[28] D. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, pp. 1–32, 2000.
[29] A. Khotanzad and Y. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, May 1990.
[30] W.-Y. Kim and Y.-S. Kim, "A region-based shape descriptor using Zernike moments," *Signal Process., Image Commun.*, vol. 16, no. 1, pp. 95–102, 2000.
[31] S. Li, M.-C. Lee, and C.-M. Pun, "Complex Zernike moments features for shape-based image retrieval," *IEEE Trans. Syst. Man Cybern. A, Syst. Humans*, vol. 39, no. 1, pp. 227–237, Jan. 2009.
[32] O. Pizarro and H. Singh, "Toward large-area mosaicing for underwater scientific applications," *IEEE J. Ocean. Eng.*, vol. 28, no. 4, pp. 651–672, Oct. 2003.
[33] N. Kumar, A. Lammert, B. Englot, F. Hover, and S. Narayanan, "Directional descriptors using Zernike moment phases for object orientation estimation in underwater sonar images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2011, pp. 1025–1028.
[34] M. Hall *et al.*, "The weka data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.
[35] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B (Methodological)*, vol. 39, pp. 1–38, 1977.

Naveen Kumar (S'10) was born in Kolkata, India, in 1987. He received the B.Tech. degree in instrumentation engineering from the Indian Institute of Technology, Kharagpur, India, in 2009. He is currently working toward the Ph.D. degree at the Department of Electrical Engineering, University of Southern California (USC), Los Angeles, CA, USA.

His research interests include machine learning and signal/image processing for applications to speech and multimedia problems.

Mr. Kumar was awarded the Viterbi School Dean's Doctoral Fellowship at USC in 2009. He has also worked on teams that have won the Interspeech 2012 Speaker Trait Challenge and the Northern Digital Inc. Excellence Awards at the 2014 International Seminar on Speech Production (ISSP).

Urbashi Mitra (S'88–M'98–SM'04–F'07) received the B.S. and M.S. degrees in electrical engineering and computer science from the University of California at Berkeley, Berkeley, CA, USA, in 1989 and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 1994.

After a six-year stint at the Ohio State University, Columbus, OH, USA, she joined the Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA, where she is currently a Professor. She has held visiting appointments at the Delft University of Technology, Stanford University, Rice University, and the Eurecom Institute. She served as Co-Director of the Communication Sciences Institute at the University of Southern California from 2004 to 2007.

Prof. Mitra is a member of the IEEE Information Theory Society's Board of Governors (2002–2007, 2012–2014) and the IEEE Signal Processing Society's Technical Committee on Signal Processing for Communications and Networks (2012–2014). She is the recipient of the 2012 Globecom Signal Processing for Communications Symposium Best Paper Award; the 2012 NAE Lillian Gilbreth Lectureship, USC Center for Excellence in Research Fellowship (2010–2013); the 2009 DCOSS Applications & Systems Best Paper Award; Texas Instruments Visiting Professor (Fall 2002, Rice University); the 2001 Okawa Foundation Award; the 2000 OSU College of Engineering Lumley Award for Research; the 1997 OSU College of Engineering MacQuigg Award for Teaching; and a 1996 National Science Foundation (NSF) CAREER Award. She currently serves on the IEEE Fourier Award for Signal Processing committee and the IEEE James H. Mulligan, Jr. Education Medal committee. She has been/is an Associate Editor for the following publications: the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2012–present), the IEEE TRANSACTIONS ON INFORMATION THEORY (2007–2011), the IEEE JOURNAL OF OCEANIC ENGINEERING (2006–2011), and the IEEE TRANSACTIONS ON COMMUNICATIONS (1996–2001). She has cochaired (technical program) the 2014 IEEE International Symposium on Information Theory in Honolulu, HI, USA; the 2014 IEEE Information Theory Workshop in Hobart, Tasmania; the IEEE 2012 International Conference on Signal Processing and Communications, Bangalore India; and the IEEE Communication Theory Symposium at ICC 2003 in Anchorage, AK, USA. She was a General Cochair for the first ACM Workshop on Underwater Networks at Mobicom 2006, Los Angeles, CA, USA. She was the Tutorials Chair for the 2007 IEEE International Symposium on Information Theory, Nice, France and the Finance Chair for the 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, NV, USA.

Shrikanth S. Narayanan (S'88–M'95–SM'02–F'09) received the M.S., Engineer, and Ph.D. degrees in electrical engineering from the University of California Los Angeles, Los Angeles, CA, USA, in 1990, 1992, and 1995, respectively

He is Andrew J. Viterbi Professor of Engineering at the University of Southern California (USC), Los Angeles, CA, USA and holds appointments as Professor of Electrical Engineering, Computer Science, Linguistics and Psychology and as the founding director of the Ming Hsieh Institute. Prior to USC, he was with AT&T Bell Labs and AT&T Research from 1995 to 2000. At USC, he directs the Signal Analysis and Interpretation Laboratory (SAIL). He has published over 600 papers and has been granted 16 U.S. patents. His research focuses on human-centered signal and information processing and systems modeling with an interdisciplinary emphasis on speech, audio, language, multimodal and biomedical problems, and applications with direct societal relevance.

Prof. Narayanan is a Fellow of the Acoustical Society of America and the American Association for the Advancement of Science (AAAS) and a member of Tau Beta Pi, Phi Kappa Phi, and Eta Kappa Nu. He is also an Editor for the *Computer Speech and Language Journal* and an Associate Editor for the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, *APSIPA Transactions on Signal and Information Processing*, and the *Journal of the Acoustical Society of America*. He was also previously an Associate Editor of the IEEE TRANSACTIONS OF SPEECH AND AUDIO PROCESSING (2000–2004), IEEE SIGNAL PROCESSING MAGAZINE (2005–2008), and the IEEE TRANSACTIONS ON MULTIMEDIA (2008–2011). He is a recipient of a number of honors, including Best Transactions Paper awards from the IEEE Signal Processing Society in 2005 (with A. Potamianos) and in 2009 (with C. M. Lee) and selection as an IEEE Signal Processing Society Distinguished Lecturer for 2010–2011. Papers coauthored with his students have won awards at Interspeech 2013 Social Signal Challenge, Interspeech 2012 Speaker Trait Challenge, InterSpeech 2011 Speaker State Challenge, InterSpeech 2013 and 2010, InterSpeech 2009—Emotion Challenge, the 2009 IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS), the 2007 IEEE Workshop on Multimedia Signal Processing (MMSP), IEEE MMSP 2006, the 2005 International Conference on Acoustics, Speech, and Signal Processing (ICASSP), and the 2002 International Conference on Spoken Language Processing (ICSLP).