OBJECT CLASSIFICATION IN SIDESCAN SONAR IMAGES WITH SPARSE REPRESENTATION TECHNIQUES

Naveen Kumar, Qun Feng Tan, Shrikanth S. Narayanan[*]

Signal Analysis and Interpretation Laboratory(SAIL), Department of Electrical Engineering,
University of Southern California, Los Angeles, CA, USA
{komathnk,qtan}@usc.edu,shri@sipi.usc.edu

ABSTRACT

Most supervised classification approaches try to learn patterns in inter class variabilities using training samples. However in the real world, their discriminative power is often diminished, because data is seldom free from irregularities within a class. Apriori modeling of these intra class variabilities poses a challenge even in underwater sidescan sonar images that we consider for object classification in this work. Sparse representation techniques prove particularly useful in this regard because of their data driven approach to model these variabilities. Results on the NSWC sidescan sonar database suggest that sparse representation classifier with zernike magnitude features is significantly robust in the presence of these non-idealities.

***Index Terms*—** Sparse Representation, Sidescan Sonar, Zernike moment, Object classification

1. INTRODUCTION

Object classification is often riddled with problems of data sparsity like most other pattern recognition paradigms. In recent years, this has led to discriminative techniques that are gaining increasing popularity. Discriminative methods focus only on the inter class differences by directly modeling the class posterior probabilities. Hence, these methods are attractive in terms of reduced number of parameters and required training data. They work well as long as the variabilities within a class are not significant. However more often than not, this poses issues on a real dataset. Various factors beyond the control of the experiment can introduce these intra class variabilities which now compete with inter class variabilities leading to a reduction in the discriminative power of the classifier. This makes robust object classification a challenging task.

In our current target application domain of underwater sidescan sonar images, this problem is indeed quite severe. Sidescan sonar imagery can provide high resolution images of the sea floor. This makes it invaluable to applications like mine-countermeasure (MCM) where speed is a key factor. While sidescan sonars allow large portions of the seabed to be scanned at once, the objects of interest in these images are usually very difficult to detect. These objects sitting camouflaged on the sea floor, are often difficult to notice except for the faint trace of a characteristic shadow adjacent to a bright highlight region (Figure 1). Additionally sidescan sonar images are very sensitive to the grazing angle, which can make the same underwater object appear completely different depending on its surrounding scene. These and other factors like texture of the object and characteristics of the medium cause significant differences in patterns within the class.

[*]Supported via ONR grant N00014-09-1-0700

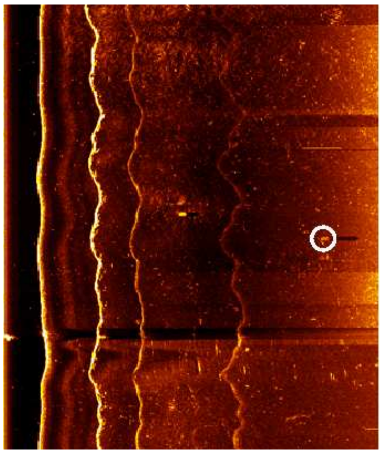

Fig. 1. A sample object (circled) in a partial sidescan sonar image with its shadow to the right.(*A colormap has been added to the original grayscale image for ease of viewing.*)

Since classical supervised approaches do not adequately perform on these datasets, there have been attempts to use knowledge specific to sidescan sonars. There has been ample work for example on extracting the object's shadow and computing features based on its shape [1]. 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 like active learning [2] and classifier fusion [3]. But in spite of such advanced methods, the false alarm rate is generally quite high. This might still be acceptable as long the false negatives are controlled. Typically detection is therefore followed by a second classification stage, where the detected object is further classified e.g. as a mine or a non-mine. Some of these works have tried to push the bar even higher by trying to further classify the detected object according to its shape [4]. If the object of interest is a mine this extra information might in some cases provide insight into ways of neutralizing the threat.

In this work we discuss an approach to solve the latter object classification problem using sparse representation methods. These methods come under the broad category of non parametric methods; this class of supervised methods does not compute any parameters or train any models from the data. This stems from their rather skeptical view that any model will have its own assumptions on the underlying distributions and hence limitations. Therefore instead of trying to compute a sufficient statistic from the features, it is safer to use the entire set of features during classification. Eccentric as it might sound, these methods are usually flexible enough to capture much of the variation in real data. Sparse representation methods specifically exploit the fact that all samples from one class are essentially distortions of a golden exemplar, and hence must lie on a lower dimensional subspace. In the context of underwater objects this would mean that features from objects of the same class in spite of their intra class differences, would share a common subspace, which can be exploited for classification. We discuss this in detail in Section 3.

A description of the dataset and the classification problem is provided in Section 2 followed by details on feature choice in Section 4. In Section 5 we discuss the different experimental procedures and their results followed by concluding points in Section 6.

2. PROBLEM DESCRIPTION

The sidescan sonar image database used for this work was collected at the Panama City Division of the National Surface Warfare Center (NSWC). The field of view in sidescan sonar images is typically to two sides of the vehicle. The dataset accordingly contains a number of pairs of left and right sonar images. Each of these 8 bit grayscale images contain one or more synthetic mine-like objects under different types of sea floor environments. Each of these objects is approximately 10-20 pixels in diameter and can be of four different types, based on the shape (Figure 2). The dataset additionally contains ground truth labels on object type and location. This work deals with the problem of this object classification, given that the object location has already been detected by one of the existing algorithms mentioned above.

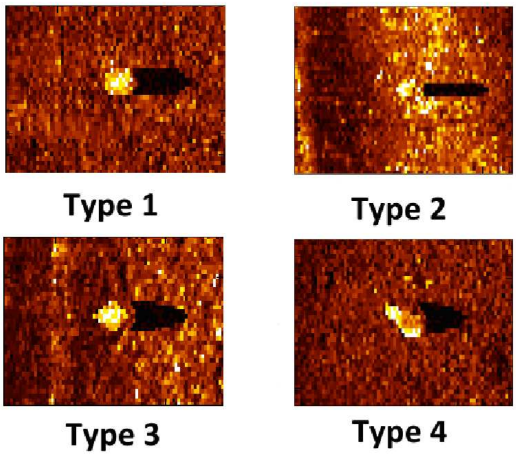

Fig. 2. The four types of object in the NSWC sidescan sonar database

3. SPARSE REPRESENTATION FOR CLASSIFICATION

The fundamental idea in this method is loosely similar to subspace models. Sparse representations as we will see, are merely a method to unmask this inherent subspace structure in the data, so that it can be exploited for classification.

3.1. Subspace Models

Subspace models assume that samples from a single class lie on a lower dimensional linear subspace. In fact given sufficient training data from each of the class it should be possible to represent any test sample as the linear combination of training samples from its class. For example if $\{x_i^{(k)}\}_{i=1}^n$ be training samples for object k and $y^{(k)}$ be a test sample from the same class then $y^{(k)}$ will approximately lie in the span of the above training samples.

$$y^{(k)} = \sum_{i=1}^{n} \alpha_i x_i^{(k)} \tag{1}$$

for some real scalars α_i

In the context of object recognition these subspace assumptions have been studied in depth for objects under various lighting and illuminations [5]. When applied to face recognition these are usually called *face subspaces* and have been shown to capture variations in expressions [6].

Given that this assumption holds, we now construct a dictionary A consisting of all training samples from all classes. For classifying a test sample y we would like to now exploit the fact that y can be represented as the sparse linear combination of a few rows of A (the ones containing samples corresponding to the object k). In other words if we write y as follows

$$y = Av \tag{2}$$

then v should be a sparse vector with only a few non zero values. To ensure sparsity of v, this can now be posed as a constrained minimization problem where we try to find the vector v with the minimum norm which also satisfies Equation 2. Typically, Equation 2 is seldom used as an equality constraint. It is usually reduced to an L2 norm of the reconstruction error which is jointly minimized (Equation 3).

Choosing the kind of norm for v is crucial here. Minimizing the L2 norm for example has been shown to return non-sparse solutions. On the other hand seeking the sparsest solution would require us to minimize the L0-norm which counts the number of non zero entries in v. This problem however is combinatorial in nature, and considered computationally infeasible. The more commonly taken middle road is that of minimizing the L1 norm. There exists some evidence that given the v is sparse enough, the solution to the L1 minimization problem approaches the solution to the L0 minimization problem [7].

3.2. Classification Methodology

For the sparse representation approach, as a baseline, we adopt the LASSO algorithm [8]. We choose this algorithm because of its prevalent usage and robust performance in most circumstances. The LASSO formulation can be written as follows:

$$\min ||y - Av||_2^2 + \lambda_1 ||v||_1 \tag{3}$$

It has however been demonstrated experimentally in speech recognition that collinearity of spectral exemplars in the dictionary leads to deterioration of the performance of LASSO [9]. An improvement can then be obtained in this respect by use of the Elastic Net algorithm [10], which better handles the situation of an incoherent dictionary. The Elastic Net algorithm is given by the following optimization formulation:

$$\min ||y - Av||_2^2 + \lambda_1 ||v||_1 + \lambda_2 ||v||_2 \tag{4}$$

where $||.||_1$ and $||.||_2$ are L1 and L2 norms respectively

We observe that Equation (4) is a more general form of the LASSO formulation in Equation (3). In particular, if we set $\lambda_2 = 0$, we obtain back the LASSO formulation. An alternative interpretation exists in terms of the priors.The LASSO can be thought of as having only a Laplacian prior, while the elastic net is a mix between a Laplacian prior and a Gaussian prior.

We adopt the Least Angle Regression (LARS) [11] implementation of both the LASSO and the Elastic Net algorithm in this paper. LARS being a greedy approach, allows us to control sparsity more efficiently.

4. FEATURE SELECTION

The role of feature extraction in pattern recognition problems is usually considered to be paramount. The general consensus seems to be that there is no *best* feature for a problem. The common approach has been to find features tuned to particular datasets, and it is no different in the domain of underwater sonar images [12].

It is interesting to see that works on sparse representation have taken a different route here. Applications to problems like face recognition have shown that given certain sparsity conditions, feature extraction ceases to play any major role in classification [6]. Given that the vector v is sparse enough (which is usually the case if the dictionary A can be made sufficiently overcomplete) even random features contain enough discriminative power to recover the correct sparse representation and hence the classification [13][14].

4.1. Polar Features

We decided to test this on two sets of features. We noticed that there still was a substantial amount of clutter around the object. Since the ground truth on location is available in the dataset and also not relevant to the classification problem at hand, we use it to compute polar domain features on images centered around the object.

The first feature simply uses raw pixel values in radial bins around the object center. We sum these pixel values in 100 uniformly quantized angle bins θ, across the radial direction for the first 10 radial pixel bins as follows.

$$P(\theta) = \sum_{\rho=1}^{10} I(\rho, \theta) \tag{5}$$

where $I(\rho, \theta)$ is the object image transformed to polar coordinates computed for angular bins $\theta = 1 \dots 100$. $P(\theta)$ is the 1-D feature which now approximately captures the shape of the object. We refer to these simply as the polar features. To make these features rotation invariant we also test shift invariant transforms of $P(\theta)$ using the circular auto correlation function.

4.2. Zernike Moments

The second feature is computed from Zernike moments of the image. Zernike moments of an image are computed via an orthogonal transform in the polar domain, where the degree of the representation controls the degree of generalizability. We use magnitudes of zernike moments which have been shown to have rotational invariance properties for object recognition [15]. Their robustness to variabilities in underwater images has also been well established [16][17]. To compute zernike moments A_{nm} of order (n, m), we find its projection with the basis function $V_{nm}(x, y)$ as follows

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{-jm\theta} \tag{6}$$

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2} - s)!(\frac{n-|m|}{2} - s)!} \rho^{n-2s} \tag{7}$$

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V^*_{nm}(\rho, \theta), x^2 + y^2 \leq 1 \tag{8}$$

where the index n is constrained to be a positive number or zero, while m can take positive or negative integer values subject to the constraints that $n - |m|$ is even and $|m| \leq n$. The range of n selects the order of zernike moments and the degree of generalizability of the description. We conduct experiments for different orders of zernike moments.

5. EXPERIMENT AND RESULTS

Once the sparse representation is computed by solving the optimization in Equation 4, classification is done using a simple heuristic. The magnitudes of coefficients are summed for each class that the corresponding atoms belong to. The test sample is classified as belonging to the class which has the maximum coefficient weight. For example in Figure 3 the classification result would be class 4 because it accumulates the maximum coefficient weight.

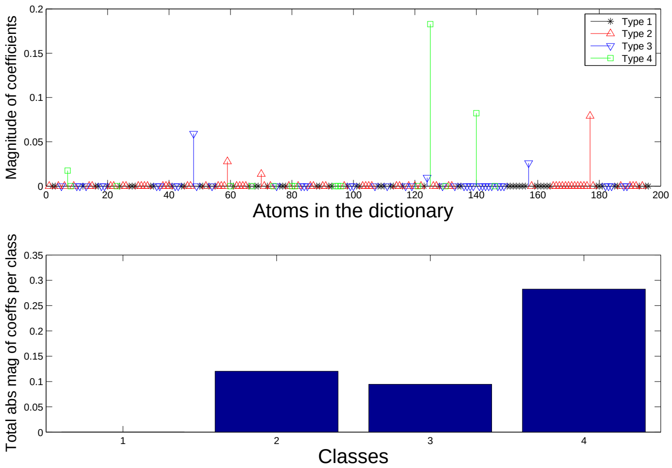

Fig. 3. Demonstration of the classification heuristic for a sample sparse representation computed on an overcomplete dictionary comprising 196 atoms.

We extract 296 objects from the dataset using the ground truth on location. The number of samples from each of the four classes was 79, 90, 83 and 44. For various comparison points we compare against standard linear classifiers like SVM, Logistic Regression and Naive Bayes using the same set of features. For the dataset of 296 objects, we perform 4-fold cross-validation. Our training data hence comprises 222 objects and test data 74 objects. WEKA [18] implementations were used for all three linear classifiers.

	NBayes	LReg	SVM	L1	E-Net
Polar	90.5	94.9	95.2	74.7	90.5
Correlation	46.3	81.4	70.3	50	72
Zernike[$n \leq 10$]	91.2	93.6	95.9	87.5	96.6

Table 1. Object classification accuracies for a 4-fold cross validation for different features and classifiers

(NBayes: Naive Bayes LReg : Logistic Regression SVM : Support Vector Machine L1 : Lasso E-Net : Elastic Net)

The results in Table 1 show that the Elastic Net is significantly more robust than LASSO for the object classification task, and is also significantly more robust than the other machine learning algorithms we tested on. This testifies to the strength of the Elastic Net in dealing with collinear feature vectors extracted from the sidescan sonar images.

We also observe that the sparse representation classifier in our current framework is in fact sensitive to the choice of features. This maybe because the dictionary was not overcomplete enough due to insufficient training data or choice of number of feature dimensions. Autocorrelations of the polar features in spite of their rotation invariance don't seem to have much discriminative power. This is most likely because they are a many to one mapping from the set of polar features. We additionally note that the optimum order of zernike moments here agrees with the observation in our previous work [17]. 36 coefficients for 10th order zernike moments capture just enough information for classification without over fitting to the samples.

6. CONCLUSION

In this paper we propose a method to deal with the intra class variabilities of objects in a sidescan sonar image for object classification. Training samples from each class form a lower dimensional subspace which is closest to the samples from that class. Sparse representation exploits this inherent structure to provide a discriminative ability without an explicit training. Results on the NSWC sidescan sonar database suggest that the method is robust in presence of variabilities.

In future work we would like include non-mine as one of the classes. This might be possible because prior work on similar out of vocabulary sparse representations, suggest that if the test sample does not belong to any of the classes in the dictionary the obtained representation will fail to be sparse. It maybe be helpful to think of a metric in that case to measure the degree of sparsity of v. Similar extension to metrics to measure the confidence of classification should also be interesting in this case.

7. REFERENCES

[1] 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. CAD/CAC Conf*, p. 2001.

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

[3] G.J. Dobeck, "Algorithm fusion for automated sea mine detection and classification," in *OCEANS, 2001. MTS/IEEE Conference and Exhibition*. IEEE, 2001, vol. 1, pp. 130–134.

[4] S. Reed, Y. Petillot, and J. Bell, "Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information," in *Radar, Sonar and Navigation, IEE Proceedings-*. IET, 2004, vol. 151, pp. 48–56.

[5] R. Basri and D.W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 218–233, 2003.

[6] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 210–227, 2008.

[7] D.L. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[9] Q.F. Tan, P.G. Georgiou, and S.S. Narayanan, "Enhanced sparse imputation techniques for a robust speech recognition front-end," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, pp. 2418–2429, 2011.

[10] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[12] J. Stack, "Automation for underwater mine recognition: current trends and future strategy," in *Proceedings of SPIE*, 2011, vol. 8017, p. 80170K.

[13] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *Information Theory, IEEE Transactions on*, vol. 54, no. 5, pp. 2210–2219, 2008.

[14] D.L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, pp. 1–32, 2000.

[15] A. Khotanzad and Y.H. Hong, "Invariant image recognition by zernike moments," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 5, pp. 489–497, 1990.

[16] O. Pizarro and H. Singh, "Toward large-area mosaicing for underwater scientific applications," *Oceanic Engineering, IEEE Journal of*, vol. 28, no. 4, pp. 651–672, 2003.

[17] N. Kumar, A.C. Lammert, B. Englot, F.S. Hover, and S.S. Narayanan, "Directional descriptors using zernike moment phases for object orientation estimation in underwater sonar images," .

[18] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, "The weka data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.