

Discriminative Sparsity for Sonar ATR

John D. McKay*, Raghu G. Raj†, Vishal Monga*, and Jason Isaacs‡

* Pennsylvania State University, University Park, PA

†U.S. Naval Research Laboratory, Washington DC

‡Naval Surface Warfare Center Panama City, FL

Abstract—Advancements in Sonar image capture have enabled researchers to apply sophisticated object identification algorithms in order to locate targets of interest in images such as mines [1] [2]. Despite progress in this field, modern sonar automatic target recognition (ATR) approaches lack robustness to the amount of noise one would expect in real-world scenarios, the capability to handle blurring incurred from the physics of image capture, and the ability to excel with relatively few training samples. We address these challenges by adapting modern sparsity-based techniques with dictionaries comprising of training from each class. We develop new discriminative (as opposed to generative) sparse representations which can help automatically classify targets in Sonar imaging. Using a simulated SAS data set from the Naval Surface Warfare Center (NSWC), we obtained compelling classification rates for multi-class problems even in cases with considerable noise and sparsity in training samples.

I. INTRODUCTION

The popularity of automated underwater vehicles (AUVs) for the purposes of mine identification and location has made Sonar-specific object recognition algorithms an important topic of study. The advantages in safety, maneuverability, and portability associated with AUVs vastly outweigh the benefits of manned vessels, but AUVs do put a particular emphasis on software that can parse non-threatening items from mines. Indeed, a wrong classification can be extremely costly by either allowing for a mine to remain undisturbed or by causing the AUV's operator to enact expensive mine extraction measures for false alarms.

Among the many difficulties associated with Sonar automatic target recognition (ATR), we look specifically at the complications associated with noisy and blurry image classification as well as limited training sizes and foreign object detection. Current Sonar image acquisition procedures can suffer from high noise and smearing which can significantly degrade the performance of many popular ATR algorithms. Even for the strategies that can handle those effects, Sonar images of the same object from different angles and/or orientations can present vastly different outputs. This high degree of variability of Sonar images of a given target with respect to angle and orientation can severely degrade ATR performance if not properly accounted for; and thereby puts a premium on the ability of ATR algorithms for robustly classifying targets under low training samples (conditioned on orientation and other relevant estimated parameters) lest one suffer from the

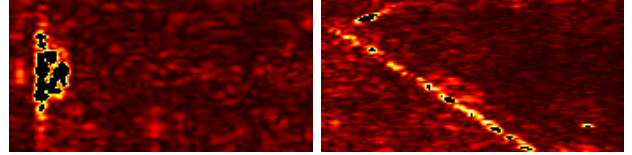


Fig. 1: Two SAS images of cylinders placed at different angles to a simulated AUV. Images courtesy of RAW SAS dataset.

complications arising from a large training set. Further, an ATR algorithm must be able to discern between that which is in its training set and foreign objects, i.e. targets that the software has not trained on. As outlined in [3] for the case of active Sonar ATR algorithms (which are designed to include human input), it is important for ATR approaches to include the capability to detect new objects. The inability to do so leads to false alarms servicing which can be very expensive in operational Sonar ATR systems. The difficulty of detecting out of class samples is further accentuated by the fact that, as pointed out above, Sonar imagery typically display high degree of translational variance (with respect to the position and viewing angle of the sensor) which together with the effects of noise and blur tend to greatly increase the variability within each class (which in turn is generally related to increase training set sizes). All these factors make Sonar ATR a complicated and highly challenging problem.

The driving method that we employ to address some of the outstanding challenges described about is derived from the work of Wright *et al* [4]. In their seminal work, the authors demonstrated an ability to produce state-of-the-art classification rates on facial images – despite immense noise and occlusion—with their implementation of a sparse-reconstruction based classification method (SRC), a creative adaptation of modern sparsity-based image compression concepts. Since its publication, researchers from many different fields ranging from biology to radar have found success in applying SRC algorithms with its powerful robustness to distortions. For Sonar, [5] demonstrated that the SRC approach can yield better accuracy results than many popular classification strategies like elastic-net or logistic regression. In this paper we take the next step and present the application the work of Wright *et al* to noisy and blurry Sonar images while also examining how well the SRC approach performed with limited training samples. Additionally, we look at how well

This work is supported by the Office of Naval Research, Arlington, VA under Grant 0401531.

the sparse concentration index (SCI) metric outlined in [4] worked in discerning objects in our training set and foreign items that could be misclassified.

In the next section, we present the mathematical underpinnings to the SRC approach, including the ℓ_1 minimization technique that we chose for our experiments. In Section 3 we demonstrate the performance of the SRC algorithm on the RAW SAS dataset obtained from the Naval Surface Warfare Center (NSWC). To give context, we provide the results of a competing algorithm based on SIFT features and compare our SRC method to this. Lastly, we conclude in Section 4 with a discussion of our results and directions of the future research.

II. METHODS

In the field of compressive sensing, which looks to represent images from relatively few samples, sparsity constraints have been used with linear models to perform impressive feats of denoising and deblurring [6]. The formulation of image reconstruction via sparsity based constraints is as follows: let $\mathbf{y}$ be the vectorized samples of noisy image measurements, A an overcomplete dictionary, and $\mathbf{x}$ the coefficients vector such that $A\mathbf{x}$ represents the vectorized reconstructed image. The idea is to minimize the number of nonzero elements in $\mathbf{x}$ while constraining the least squares error of the reconstructed image within some tolerance $\varepsilon > 0$. That is, we look to solve

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{y} - A\mathbf{x}\|_2 < \varepsilon \quad (1)$$

where $\|\cdot\|_0$ is the ℓ_0 quasi-norm [7]. Wright *et al* looked to adapt this same strategy to classification problems [4]. In their case, the problem is to figure out which class the vector $\mathbf{y}$ belonged to. To this end, they assigned the columns of matrix A to be vectorized training images corresponding to each class; i.e. if we let A_j be a matrix whose columns each correspond to a vectorized training image of class j , then $A = [A_1 \ A_2 \ \dots]$. They were then able to find the class assignment for $\mathbf{y}$ by solving (1) and attributing the identity of the test image to the class that yields the smallest residual error. In other words, for $\delta_j(\mathbf{x})$ defined to be a characteristic function that outputs a vector consisting of the indexes associated with class j and zeros everywhere else, they would identify the class of $\mathbf{y}$ by solving

$$\min_j \|\mathbf{y} - A\delta_j(\mathbf{x})\|_2$$

It is widely known in compressive sensing that the solution of (1) is a NP-hard problem. The issue lies in the ℓ_0 quasi-norm which renders the optimization problem to be highly non-convex and whose solution amounts to a computational intractable combinatorial search. Because of this, it is common practice to relax the ℓ_0 quasi-norm to its best convex approximation, the ℓ_1 norm. The key to this relaxation is the fact that the resulting mathematical program still promotes sparsity under suitable incoherence conditions of dictionary A . Thus, instead of (1), we focus on solving

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y} - A\mathbf{x}\|_2 < \varepsilon \quad (2)$$

and use the same strategy as before to determine the classification.

There are many different algorithms available to solve (2). [8] performed a comprehensive comparison of many different ℓ_1 minimization approaches and found consistently that Homotopy methods worked best in denoising and classifying applications. Based on this and other tests we conducted, we focus on Homotopy methods as our primary optimization algorithm for our SRC process. One can find detailed descriptions of Homotopy approaches in [8] and [9], but in short these methods use the subgradient of the objective function

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2 + \lambda \|\mathbf{x}\|_1$$

where $\lambda \in \mathbb{R}$ represents a sparsity-enforcement term, and the fact that there exists a homotopy (continuous deformation) between $F(\mathbf{x})$ and a similar objective function with a two norm relaxation of the ℓ_0 quasi-norm as λ approaches zero.

III. EXPERIMENTS

To test the SRC approach in Sonar image classification, we used the RAW SAS database obtained from the NSWC. RAW SAS contains 13 sets of object configurations, each one with a specific arrangement of 4 objects against 11 different background settings. While backgrounds were obtained from actual synthetic aperture Sonar measurements, the objects were realistically simulated according to their location, angles to the sensors, etc and superimposed on the scene. Each object was snipped from its scene and saved as a target chip; this chip included possible background interference found within the rectangular window used to collect the sample. The objects consisted of blocks, cones, cylinders, spheres, toruses, and pipes, though because the latter two were meagerly represented we tested primarily on the first four (main classes) and leave the last two, pipes and toruses (foreign classes), for our tests on foreign object detection.

Each of our tests had a similar experimental design. We crafted 20 different dictionaries out of an equal number of samples from each class and tested them against 40 test images coming from 10 remaining samples per main class. For the tests of sample size, we varied the size of the dictionaries to see how well they could classify the 40 images, while for noise and blur we kept the dictionary size constant at 100, that is, 25 from each main class. Gaussian white noise was added at an increasing noise variance to each of the testing sets while the blurring was added at an increasing Gaussian filter.

Object	Samples
Blocks	88
Cones	66
Cylinders	308
Spheres	66
Pipes	22
Toruses	22

TABLE I: Number of each object in our dataset.

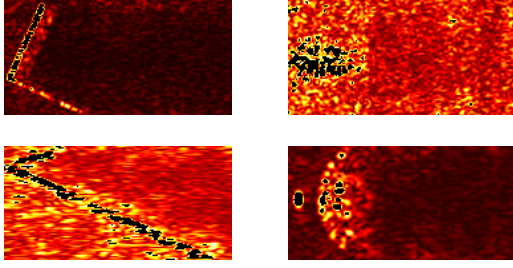


Fig. 2: Examples of target snippets from RAW SAS; top left: block, top right: cone, bottom left: cylinder, and bottom right: sphere

To provide context, we also used an approach inspired by [10] where the authors demonstrated that a SIFT-based SVM classifier can handle many of the nuances underlying Sonar image classification quite well. For 20 trials, we took a certain number elements from each main class as our the training set and extracted their SIFT features. We then crafted a codebook by clustering these SIFT features and found the corresponding histogram representation of all the training images. Next, we trained an SVM based on the histogram representations and selected a testing set of 40 images (10 per main class) to apply to this classifier. We then took each of the test images and converted all of them to their histogram representation corresponding to the codebook which ultimately allowed us to classify them with the trained SVM. When it came to testing the performance under different training sample sizes, we varied the SIFT SVM's training equally to the SRC's. As for noise and blur, we tested the SIFT SVM approach with a training set equal in size to the SRC, 25 samples per main class, and with a size much greater than the SRC's, 40 per main class.

A. Training Sample Size

First we looked to address the performance of the SRC approach given limited training set sizes. Figure 3 provides a comparison between the SRC and SIFT SVM approaches over various training. Note that by 20 elements from each class - which represents less than a third of even the smallest class - our SRC dictionaries were able to obtain better than 95% accuracy with their 40 test images. On the other hand, the SIFT SVM approach was unable to yield an average accuracy rate above 90% until 35 samples from each class and was vastly outperformed by the SRC method for the smaller training sizes.

With regards to the classification rate of each individual classes, Figure 4 shows on average how well they were identified over the different training sizes. The block and sphere cases stand out as examples wherein the SRC algorithm was able to perform at a high rate even in limited training settings - especially the spheres where even with as few as 5 training samples, the SRC provided classification rates above 95%. The cones seemed to give neither algorithm much of a challenge while the cylinders proved to be an interesting

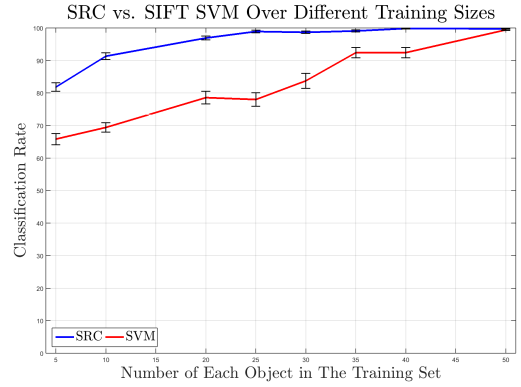


Fig. 3: Classification rates over various training set sizes.

case in variability. The many different angles and orientations that we had for the cylinders gave the SIFT SVM trouble in providing predicable results, while the SRC was still able to demonstrate the monotonically increasing classification rate pattern we would expect in this experiment. This speaks to the potential reliability the SRC algorithm may offer that Sonar ATR needs: predictability.

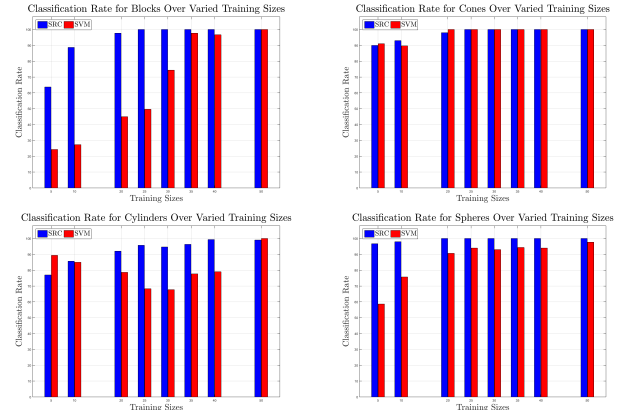


Fig. 4: Class specific classification rates over varied training.

B. Classification Under Noise

Like with limited training samples, Wright *et al* showed how well the SRC approach can handle noise. In Sonar applications where a misclassification can be costly, the onus is on the software to demonstrate a resiliency to noise. In is in this setting that we performed our next set of experiments on the RAW-SAS dataset.

In every case the SRC method outperformed the SIFT SVM approach. As demonstrated in Figure 5, the SRC approach was able to provide much higher classification rates than the SIFT SVM in equal training environments and was even able to slightly edge the SIFT SVM with many more training samples in all but the heaviest noise scenarios. While this confirmed our suspicions given how successful Wright *et al* were with their facial databases, it is still impressive how well the SRC approach handles Sonar images especially considering the

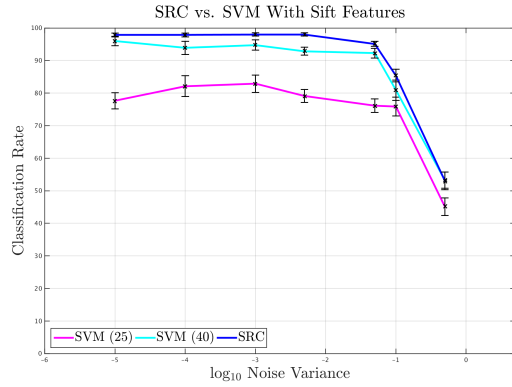


Fig. 5: Noisy image classification rates with standard error bars.

presence of significant background clutter already present even without the noise. That said, the scenes with the highest noise did give our SRC approach trouble, though not as much as the SIFT SVM with equal training.

Figure 6 gives an idea of how each main class faired with the SRC and SIFT SVM approach with equal training. These image depict how the SIFT SVM had problems deciphering between spheres and cones; as 2 shows, they do share similar geometric characteristics making for the possibility of difficulty when it came to parsing between them with SIFT features. The SRC method, on the other hand, did not appear to have trouble with any one class and, instead, saw a degradation in classification rates overall.

It is worth noting how the SRC approach was able to rightly identify images 95% of the time or better up until the noise reached a variance of .1 and images with approximate SNR values of $-10dB$. Therefore, given the relative performance, the SRC demonstrated promise but still has trouble under heavy noise.

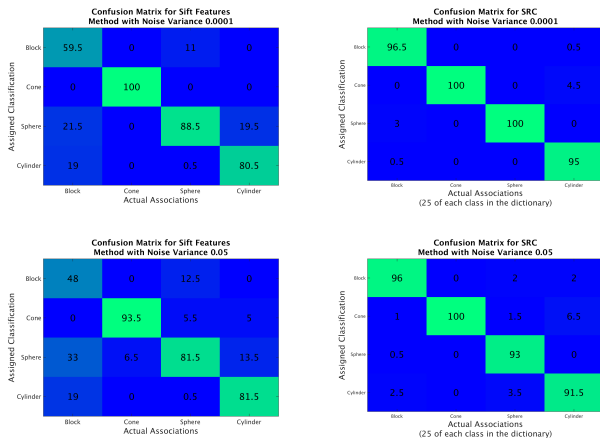


Fig. 6: Scaled confusion matrices for lower and higher variance noise for the SRC and SIFT SVM with equal training. The columns represent the actual tested image class and the rows their classified class (sums of columns are 100%).

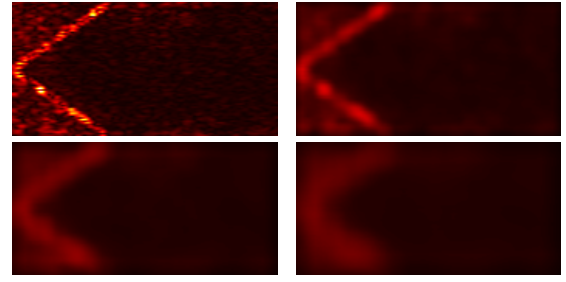


Fig. 7: Example blurring on a block; top left is the original, bottom left is a lower level blur, top right is a higher blur intensity, and bottom right is the highest.

C. Classification Under Blurring

Much like noise, AUV systems must be ready to handle non-trivially blurred Sonar images. We tested our SRC algorithm and the SIFT SVM approach with increasingly blurry images using simple Gaussian filters on test images. We use the term “blurring intensity” to refer to the value $b \in \mathbb{R}$ that represents the pixel dimension of the blurring as well as the Gaussian standard deviation for the filter. To get an idea of how the blurring progressed in our experiment, Figure 7 gives an example of blurred test images.

The SRC approach was able to sustain better classification rates over higher intensity blur than the popular SIFT SVM method in our tests. In fact, the SRC approach was able to vastly outperform the SIFT SVM method with increased blurring to a degree greater than with the noise, as Figure 8 demonstrates. As we see, as blurring reach a certain severity, the added training became irrelevant for the SIFT SVM approach, whereas the SRC was still able to produce usable classification rates. With regards to each class, the SRC method showed resiliency in all categories but with spheres, which turned out to be the class that gave the method the most challenge. The blur did little to impact the classification of cones and cylinders, though under heavy blurring, the SRC approach tended to confuse blocks and spheres with cylinders. That said, the SIFT SVM approach appeared to fall victim to the rounding type effect heavy blurring had on tests and had difficulty discerning blocks, cones, and cylinders from spheres. Strategies that rely heavily on the shape of the object, such as SIFT feature-based algorithms, may be more susceptible to these types of errors. Figure 9 provides the confusion matrices for our blur experiment.

Despite the troubles, given that the SRC approach proved to have an accuracy rate higher than 60% even in the most intense blurring examples we tried, we see that this strategy may hold a great deal of promise. In times where SAS imaging fails to produce crisp images, the SRC algorithm can still provide reliable classification rates without the troubles associated with approaches such as the SIFT SVM.

D. Foreign Object Detection

In [3], the authors outlined an active Sonar image classification system that had a built in method to find new

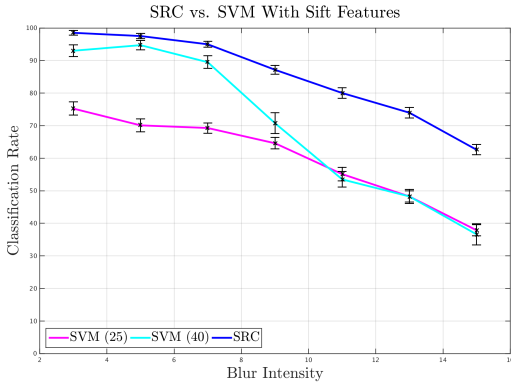


Fig. 8: Blurring experiment results for SRC and SIFT SVM with standard error bars.

objects not found in the training dictionary. We wished to see how the sparsity concentration index (SCI) outlined in Wright *et al* could perform in a similar task of identifying images outside the training set and, importantly, keeping them from automatically being misclassified into one of the predetermined static number of classes. The SCI metric for a coefficient vector $\mathbf{x}$ found through a SRC approach is given by:

$$\text{SCI}(\mathbf{x}) = \frac{(\max_i \|\delta_i(\mathbf{x})\|_1 / \|\mathbf{x}\|_1) k - 1}{k - 1}$$

The key to this metric is that it tends to 1 if $\mathbf{x}$ is sparse and to 0 otherwise. Since SRC approaches attempt to force $\mathbf{x}$ to be sparse based on a constructed dictionary A , if a test vector is foreign with respect to the elements of A , then there is a good chance that $\mathbf{x}$ we not be very sparse. Therefore, we can decide to reject a test vector's classification if its SCI value is below a threshold $\kappa \in (0, 1)$.

To test the SCI metric we used the two smaller classes from the RAW SAS dataset, the pipes and toruses which we call

Object	SCI with Training Sizes of:		
	25	35	45
Block	.478 (.130)	.463 (.106)	.683 (.151)
Cone	.169 (.059)	.175 (.051)	.471 (.154)
Cylinders	.501 (.088)	.438 (.079)	.745 (.089)
Spheres	.184 (.062)	.181 (.056)	.466 (.182)
Torus	.173 (.038)	.201 (.033)	.195 (.053)
Pipe	.193 (.054)	.216 (.046)	.329 (.097)

TABLE II: Average SCI values with standard deviations for our tests on dictionaries that used 25, 35, and then 45 elements from each main class.

the foreign classes, and tested them against dictionaries built with only the main classes. We varied the dictionary sizes between 25, 35, and 45 members of each main class as to get a better idea of how the training sample can influence this metric. The tests consisted of 20 remaining elements of the main classes and 10 from both of the foreign classes and we found by the SCI and the assigned classes. With these, we selected several threshold values κ and evaluated how well that value discerned the foreign objects from the main ones as well as how this influenced the overall classification performance of the exercise.

We found that the SCI is able to distinguish to some extent between the objects in the dictionary and the newly introduced torus and pipe. As Table II shows, the SCI metric did its best discriminating between the main and foreign classes when the dictionaries were at their largest. The lower SCI values for cones and spheres turned out to be an interesting finding and could potentially be due to their smaller are on the target chips, though with larger dictionaries, they were able to differentiate themselves from the foreign classes on average. When looking at the classification rates, as Figure 10 demonstrates with dictionaries containing 45 elements from each main class, we see some of the problems with using the SCI to weed out foreign objects. For example, while the pipes had on average a lower SCI value, their variance was less than that of the cones and spheres, so when we increased the threshold κ to .15 and .25, we saw many correctly classified cones and spheres removed. At that same time, even with the highest thresholds, the blocks and cylinders were barely effected. Therefore, the SCI metric may be useful in certain cases such as with our dataset if we were concerned with finding only blocks and cylinders, but more investigation must be done to refine the SCI so it can become a tool in foreign object detection.

IV. DISCUSSIONS AND FURTHER WORK

In this paper we have applied the SRC methodology to Sonar ATR and found robust performance in varying amounts of noise and blur. Furthermore this method continues to exhibit superior performance to the popular SIFT SVM based approach in limited training size regimes. It is well known however that the performance of SRC is sensitive to the correctness of pose estimation in both the training and test

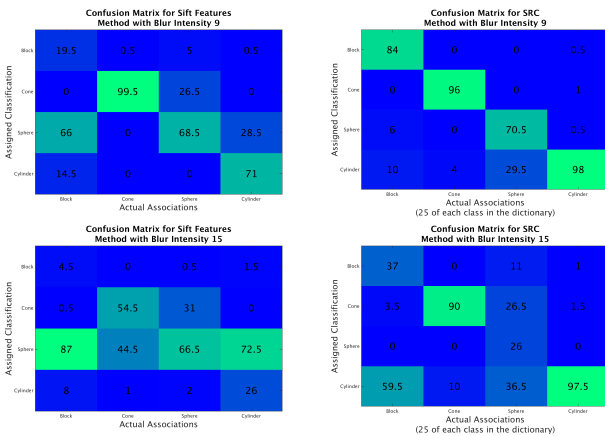


Fig. 9: Scaled confusion matrices for mid and high level blurring with SRC and SIFT SVM under the equal training. The columns represent the actual tested image class and the rows their classified class (sums of columns are 100%).

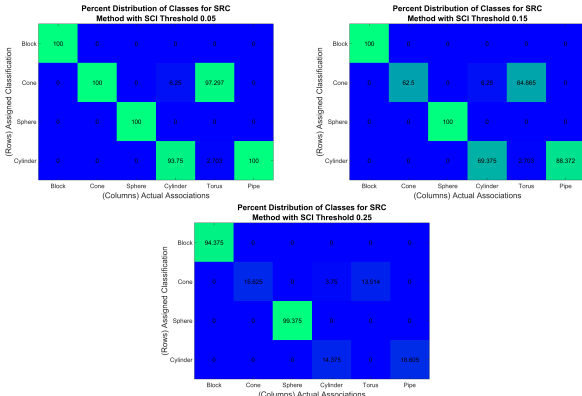


Fig. 10: Classification rate over different κ of dictionaries with 45 elements from each main class.

samples. Given this an important next step would be to examine the problem if image pose.

The peculiarities of the Sonar imaging modality, however, demands solutions that are somewhat different from those that are successful for optical images (in particular for facial images). In our experiments we carefully focused on Sonar images obtained from a normal side-scan capture. More generally, the angles at which an object faces the Sonar device greatly influences its appearance inasmuch as a slight adjustment in its position can render vastly different image projections. Given this we posit that it would be prudent to focus on translation invariant methods when extracting image features conditioned on the pose of the object under consideration. This translational invariance property would also enable the classification of images with inexact window sizes. Our future investigations aim to resolve these issues while also incorporating powerful Bayesian priors on the underlying dictionary structure [11] [12] [13] to enable robust classification performance in increasingly difficult clutter environments.

REFERENCES

- [1] T. Fei, D. Kraus, and A. Zoubir, "Contributions to automatic target recognition systems for underwater mine classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 1, pp. 505–518, Jan 2015.
- [2] R. Fandos, L. Sadamori, and A. Zoubir, "Sparse representation based classification for mine hunting using synthetic aperture sonar," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 3393–3396.
- [3] E. Kriminger, J. Cobb, and J. Principe, "Online active learning for automatic target recognition," *Oceanic Engineering, IEEE Journal of*, vol. 40, no. 3, pp. 583–591, July 2015.
- [4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.
- [5] N. Kumar, Q. F. Tan, and S. S. Narayanan, "Object classification in sidescan sonar images with sparse representation techniques," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1333–1336.
- [6] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

- [7] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [8] A. Y. Yang, S. S. Sastry, A. Ganesh, and Y. Ma, "Fast ℓ_1 -minimization algorithms and an application in robust face recognition: A review," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 1849–1852.
- [9] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Homotopy continuation for sparse signal representation," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 5. IEEE, 2005, pp. v–733.
- [10] Z. Zhu, X. Xu, L. Yang, H. Yan, S. Peng, and J. Xu, "A model-based sonar image atr method based on sift features," in *OCEANS 2014 - TAIPEI*, April 2014, pp. 1–4.
- [11] U. Srinivas, Y. Suo, M. Dao, V. Monga, and T. D. Tran, "Structured sparse priors for image classification," *Image Processing, IEEE Transactions on*, vol. 24, pp. 1763–1775, June 2015.
- [12] U. Srinivas, V. Monga, and R. G. Raj, "SAR automatic target recognition using discriminative graphical models," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 50, pp. 591–606, January 2014.
- [13] R. G. Raj and A. Bovik, "A hierarchical bayesian-map approach to computational imaging," *IEEE International Conference on Image Processing*, 2014.