SPARSE REPRESENTATION BASED CLASSIFICATION FOR MINE HUNTING USING SYNTHETIC APERTURE SONAR

Raquel Fandos, Leyna Sadamori and Abdelhak M. Zoubir

Signal Processing Group, Institute of Telecommunications
Technische Universität Darmstadt
Merckstr. 25, 64283 Darmstadt, Germany
Phone Number: +49 6151-16 70804, Fax Number: +49 6151-16 3778
E-mail: {rfandos, sadamori, zoubir}@spg.tu-darmstadt.de

ABSTRACT

In this paper, a Sparse Representation based Classification (SRC) approach is employed for mine hunting using Synthetic Aperture Sonar (SAS) images. Given a training database with enough samples, SRC exploits the properties of sparse signals and expresses a sample of unknown class as a sparse linear combination of the training samples. The class of the training samples with greater weight is likely to be the candidate sample class. The method was introduced for face recognition, where the face images are directly taken as feature sets. Due to the greater variability of sonar images, for mine hunting applications it is more convenient to transform the image samples into a different feature domain. Several feature sets are considered, and the results are compared with those provided by a linear discriminant analysis classifier. We have tested the method on an extensive SAS database with more than 400 mines.

Index Terms— sparse representation, classification, mine hunting, synthetic aperture sonar

1. INTRODUCTION

Compressive sampling has received a lot of attention in the last years [1, 2]. Exploiting the properties of sparse representation, it yields compression rates far above the Nyquist threshold. In this paper, signal sparse representation is employed for a different purpose, namely, classification. Sparse Representation based Classification (SRC) was introduced for face recognition in [3]. Given a training database with enough samples of known classes, SRC predicts the class of a candidate sample by expressing its feature set as a sparse linear combination of the feature vectors of the training samples. Subsequently, the distance between the candidate sample and the reconstructed sample stemming from each single class is calculated. The candidate sample is assigned to the class minimizing that distance.

In this paper, we employ the same principle for a different application, mine hunting based on Synthetic Aperture Sonar (SAS) images. The Automatic Detection and Classification (ADAC) of underwater objects for mine hunting is an active field of research for several decades [4, 5, 6, 7]. Fig. 1 includes a collection of SAS snapshots with several cylindrical and spherical mines. Note that the shadow of the objects are, indeed, more prominent than their highlights. The ADAC system employed in this paper consists of three stages: segmentation, feature extraction and classification. First, the objects present in the image are detected by means of a segmentation algorithm. Not only the mines, but also physical features of the terrain such as rocks and sand ripples, are segmented. A set of significant features is extracted from each object, e.g., geometrical or statistical descriptors [8]. The final task of the system is to assign one class to each object. The classes are normally 'mine' and 'no mine'. It is also possible to specify different classes for different mine types, such as 'spherical mine' and 'cylindrical mine'. The feature vector of each detected object is compared with those of a training database, whose classes are known *a priori*. The object is classified accordingly. Those mines assigned the 'no mine' class are missed and those terrain features assigned the 'mine' class constitute false alarms. The main goal of the system is the minimization of the missed detection rate while keeping a reasonably low false alarm rate.

All system stages (segmentation, feature extraction and classification) influence the final performance. In this paper we focus on the latter, and we propose classification via sparse representation as a novel classification method for mine hunting applications. The system performance is compared with that of Linear Discriminant Analysis (LDA) classifier. Different sets of features are considered.

A 57,000 m^2 database of SAS images, with a resolution of 2.5 cm $\times$ 2.5 cm per pixel, has been employed to test the system. It includes more than 400 mines, both cylindrical and spherical. We have employed a Markovian segmentation algorithm [8] to segment the images. Three regions have been considered: highlight, shadow and background (see Fig. 1 for an illustration). More than 3000 'no mine' objects are seg-

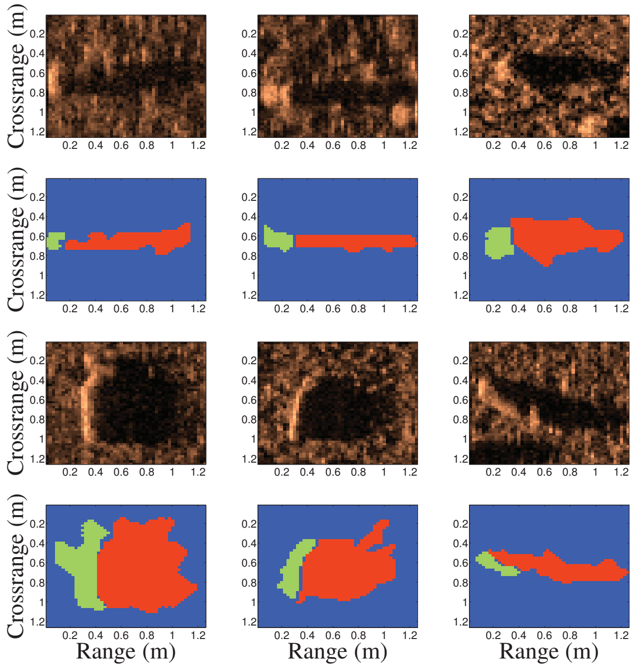

Fig. 1. The first row shows three snapshots of spherical mines. The second row illustrates their segmentation into three regions: highlight (green), shadow (red) and background (blue). the third row depicts three cylindrical mines and the forth row the corresponding segmentation results.

mented together with the man made objects of interest.

This paper is organized as follows. Sec. 2 is devoted to SRC theory. The adaptation of the method to mine hunting is tackled in Sec. 2.1. Results are presented in Sec. 3 and the paper concludes with Sec. 4.

2. SPARSE REPRESENTATION BASED CLASSIFICATION

In SRC, we arrange the feature vectors of the training samples of class k in a matrix $A_k = [\mathbf{a}_{k,1}, \mathbf{a}_{k,2}, \ldots, \mathbf{a}_{k,N_k}] \in \mathbb{R}^M \times \mathbb{R}^{N_k}$, where $\mathbf{a}_{k,n} \in \mathbb{R}^M$ denotes the feature vector of the nth sample of the kth class, N_k is the number of samples of class k, $k \in \{1, \ldots, K\}$, and K is the number of classes.

SRC was originally proposed for face recognition [3]. For such application, each class k corresponds to a different person. The database in [3] employs 'passport pictures', where the persons are always shown from the front and the poses are uniform. Therefore, it is not necessary to transform the samples into any feature space to improve the image characterization, and each feature vector $\mathbf{a}_{k,n}$ corresponds directly to the vectorized version of the image normalized in size.

Given sufficient training samples, any test sample $\mathbf{y}$ of class k can be expressed by a linear combination of the training samples of that class:

$$\mathbf{y} = \sum_{n=1}^{N_k} \mathbf{a}_{k,n} \lambda_{k,n} \quad \text{or} \quad \mathbf{y} = A_k \Lambda_k, \tag{1}$$

where the vector $\Lambda_k = [\lambda_{k,1}, \ldots, \lambda_{k,N_k}]^T$ contains the weighting coefficients.

The membership of $\mathbf{y}$ to the kth class is however unknown and needs to be predicted. For this purpose, we define a new matrix A of size $M \times N$, $N = N_1 + N_2 + \ldots + N_k$, containing the training samples of all K classes, $A = [A_1, A_2, \ldots, A_K]$. Again, we can express the test sample in terms of a linear representation:

$$\mathbf{y} = A\mathbf{s}, \tag{2}$$

where $\mathbf{s} = [0, \ldots, 0, \lambda_{k,1}, \ldots, \lambda_{k,N_k}, 0, \ldots, 0]^T \in \mathbb{R}^N$ is a coefficient vector whose entries are zero except those associated with the kth class.

From compressive sampling theory [1, 2], we know that we can find a solution $\mathbf{s}$ to the linear Eq. (2), if $\mathbf{s}$ is sparse enough, by minimizing the ℓ_1-norm of all possible solutions:

$$\hat{\mathbf{s}} = \arg\min ||\mathbf{s}||_1 \qquad \text{s.t.} \quad A\mathbf{s} = \mathbf{y}. \tag{3}$$

Ideally, the non-zero entries of $\hat{\mathbf{s}}$ will all be associated with the columns of A containing the training samples of its class k. However, due to noise and modelling errors, the estimate $\hat{\mathbf{s}}$ may have non-zero coefficients that are associated to other classes. Nevertheless, their magnitude should be small in relation to those of class k. Taking this into account, the following procedure is established in order to predict the class k of a certain test sample $\mathbf{y}$. We define a characteristic function $\delta_k \colon \mathbb{R}^N \to \mathbb{R}^N$ of the kth class, where $\delta_k(\hat{\mathbf{s}})$ sets all coefficients of $\hat{\mathbf{s}}$, which do not belong to the kth class, to zero. The residuals can now be calculated as

$$r_k = ||\mathbf{y} - A\delta_k(\hat{\mathbf{s}})||_2, \tag{4}$$

which is the ℓ_2-distance between the test sample $\mathbf{y}$ and a reconstructed sample generated by a linear combination of training samples only of the kth class. We assign class k to the test sample $\mathbf{y}$ if r_k is the minimum residual, $\forall k \in \{1, \ldots, N\}$.

Note that the images $\mathbf{a}_{k,n}$ in A need to be normalized to have unit ℓ_2-norm. This can be seen as an equalization of the image energy. Illustratively, consider a set of dark images and one bright image as the training data. If we want to classify a bright test sample, the bright training sample will dominate regardless of which class it belongs to, unless the energy normalization has been applied beforehand.

2.1. Application to Mine Hunting

As referred above, [3] employs face images directly as feature vectors $\mathbf{a}_{k,n}$. This is possible due to the nature of the images, which are uniform in size and pose, presenting only

Feature Set	Name
SAS Image	$\text{Image}^{\text{SAS}}$
Segmented Image	$\text{Image}^{\text{Seg}}$
Fourier Coefficients	FC
Principal Components	PC
Normalized Central Moments	NCM
Optimal feature set for LDA classifier	$f_{\text{LDA}}^{\text{opt}}$

Table 1. Feature sets

small variations within on class. We try to mimic these 'passport photo' conditions for mine hunting by centering the sonar images of the mines around the center of mass and normalizing their size. It remains, however, a high variability in the appearance of underwater mines within a single class. By contrast with the passport photos, the shape of an underwater mine and its shadow are dependent on many factors such as the orientation and the direction of illumination by the sonar system. Moreover, SAS images are of noisy nature. For an illustration, observe the snapshots in Fig. 1. While the first two spheres are similar, the third one is significantly different. The same can be seen for the cylinders. This strong variation of sonar images degrades the performance when they are used directly as feature vectors.

In order to improve the classification results, a collection of alternative feature domains have been tested. First, the segmentation result has been considered. While the background pixels are assigned the value 0, the highlight and shadow pixels are assigned the values 1 and -1, respectively. The 2D-Fourier Coefficients (FC) of the binary representation of the segmented shadow are considered as features. In total 49 coefficients are extracted (order seven in both range and cross-range direction). The first 50 Principal Components (PC) [9], extracted as well from the binary representation of the segmented shadow, have been considered. Moreover, we have tested the set of Normalized Central Moments (NCM) [10] up to order ten of the segmented shadow. Finally, a heterogeneous set of 72 features, hereafter referred to as $f_{\text{LDA}}^{\text{opt}}$, has been tested. This feature set has been found by the Sequential Forward Floating Search feature selection algorithm for the specific database considered in this paper and a Linear Discriminant Analysis (LDA) classifier [11]. The feature set consists of a subset of 2D-FC, PC and NCM combined with invariant moments and a few geometrical and statistical features specially designed to minimize their variability in poor segmentation scenarios.

3. RESULTS

SRC has been tested on a database of 57,000 m^2 SAS images, where 129 cylindrical mines and 308 spherical mines are present. Furthermore, 3604 clutter objects constitute the 'no mine' class. To avoid the predominance of the clutter class, only a subset of 400 clutter elements has been included in the training set (although allof them are employed to test the system).

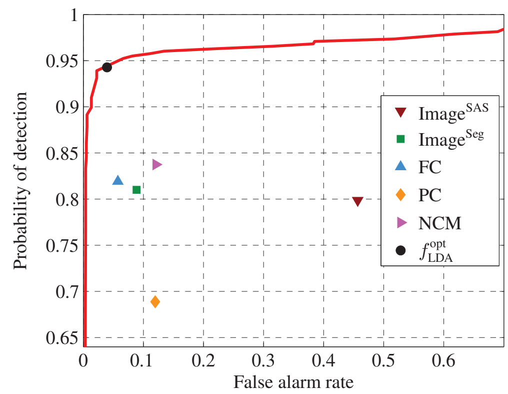

Fig. 2. SRC vs. LDA classification results

Due to the limited number of samples, all classification results are obtained using the leave-one-out technique [12]. The ℓ_1-minimization implementation is based on the *SpaRSA* algorithm proposed in [13].

SRC has been tested for the different feature sets described in Sec. 2.1. They are listed in Table 1. Results are illustrated in Fig. 2. As a reference, the ROC curve of the LDA classifier for its optimal feature set is depicted as well. Using the sonar image $\text{Image}^{\text{SAS}}$ directly as feature set produces a false alarm rate of almost 0.5 for a detection rate of 0.8. These results are well below the performance of the LDA algorithm. If the segmentation result, $\text{Image}^{\text{SAS}}$, is utilized, the detection rate slightly increases and the false alarm rate decreases to 0.1. FC and NCM yield similar classification results, all of them still not comparable to the LDA performance. Only when SRC employs the $f_{\text{LDA}}^{\text{opt}}$ feature set, its classification results are as good as those provided by LDA. For a 0.04 false alarm rate (0.0025 false alarms per m^2), almost 95 % of the mines are detected.

The computational cost of SRC has been considered. For the $f_{\text{LDA}}^{\text{opt}}$ feature set, above 1 s is required to classify each sample with an Intel i5 4 core 2.8 GHz processor. This high computational cost is due to the ℓ_1-norm calculation. By contrast, the cost of classifying one sample by the LDA classifier is about 1000 times lower.

4. CONCLUSION

Sparse representation based classification has been employed for mine hunting using synthetic aperture sonar images. By contrast with face recognition application, it is advantageous to express the sonar images in a suitable feature space. This is due to the higher variability of sonar imagery. Six different feature domains have been tested on a database of more than 400 mines. The best results have been obtained when a heterogeneous feature set, consisting of a combination of geometrical and statistical features of the segmented object, is employed. Almost 95 % of the mines are detected for a false alarm rate of 0.0025 per m^2. This performance is as good as that of the linear discriminant analysis classifier, which suggests that SRC is a powerful tool for mine hunting applications when a suitable feature domain is employed. Its computational cost is, however, significantly higher that the LDA cost.

Acknowledgment

We would like to thank Dr. A. Kraft and Dr. K. Siantidis from ATLAS ELEKTRONIK GmbH for providing the data and for their support and valuable comments.

5. REFERENCES

[1] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21 – 30, 2008.

[2] R.G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118 – 121, 2007.

[3] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210 – 227, 2009.

[4] M. F. Doherty, J. G. Landowski, P. F. Maynard, G. T. Uber, D. W. Fries, and F. H. Maltz, "Side scan sonar object classification algorithms," in *Proceedings of the 6th International Symposium on Unmanned Untethered Submersible Technology*, 1989, pp. 417 –424.

[5] G. J. Dobeck, J. C. Hyland, and L. Smedley, "Automated detection and classification of sea mines in sonar imagery," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 1997, vol. 3079, pp. 90–110.

[6] T. Aridgides, M. F. Fernandez, and G. J. Dobeck, "Side-scan sonar imagery fusion for sea mine detection and classification in very shallow water," 2001, vol. 4394, pp. 1123 – 1134, SPIE.

[7] E. Coiras, P.-Y. Mignotte, Y. Petillot, J. Bell, and K. Lebart, "Supervised target detection and classification by training on augmented reality data," *IET Radar Sonar Navigation*, vol. 1, no. 1, pp. 83 –90, 2007.

[8] R. Fandos and A. M. Zoubir, "Optimal feature set for automatic detection and classification of underwater objects in sas images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 454 – 468, 2011.

[9] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, pp. 559 – 572, 1901.

[10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[12] A. K. Jain, R. P. W. Duin, and M. Jianchang, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4 – 37, 2000.

[13] S.J. Wright, R.D. Nowak, and M.A.T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479 – 2493, 2009.