Tensor Sparsity for Classifying Low-Frequency Ultra-Wideband (UWB) SAR Imagery

Tiep H. Vu\(^1\), Lam Nguyen\(^2\), Calvin Le\(^3\), Vishal Monga\(^4\)

1,4 School of Electrical Engineering and Computer Science
The Pennsylvania State University
University Park, PA, 16802
Email: \(^1\) thv102@psu.edu
\(^4\) vmonga@engr.psu.edu

2,3 Sensors & Electron Devices Directorate
U.S. Army Research Laboratory
2800 Powder Mill Rd, Adelphi, MD, 20783
Email: \(^2\) lam.h.nguyen2.civ@mail.mil
\(^3\) calvin.d.le.civ@mail.mil

Abstract—Although a lot of progress has been made over the years, one critical challenge still facing low-frequency (UHF to L-band) ultra-wideband (UWB) synthetic aperture radar (SAR) technology is the discrimination of buried and obscured targets of interest from other natural and manmade clutter objects in the scene. The key issues are i) low-resolution SAR imagery for this frequency band, ii) targets of interests being typically small compared to the radar signal wavelengths, iii) targets having low radar cross sections (RCSs), and iv) very noisy SAR imagery (e.g. target responses buried in responses from cluttered environment). In this paper, we consider the problem of discriminating and classifying buried targets of interest (buried metal and plastic mines, 155-mm unexploded ordnance [UXO], etc.) from other natural and manmade clutter objects (soda can, rocks, etc.) in the presence of noisy responses from the rough ground surfaces for low-frequency UWB 2-D SAR images. We generalize the traditional sparse representation-based classification (SRC) to a model with capability of using the information of the shared class, and implement multichannel classification problems by exploiting structures of sparse coefficients using various techniques. Here, we employ an electromagnetic (EM) SAR database generated using the finite-difference, time-domain (FDTD) software, which is based on a full-wave computational EM method.

Keywords—SRC, sparse recovery, UWB, SAR, discriminator, classifier.

I. INTRODUCTION

Over the past two decades, the U.S. Army has been investigating the capability of low-frequency, ultra-wideband (UWB) synthetic aperture radar (SAR) systems for the detection of buried and obscured targets in various applications, such as foliage penetration [1], ground penetration [2], and sensing-through-the-wall [3]. These systems must operate in the low-frequency spectrum that spans from UHF frequency band to L band in order to achieve both resolution and penetration capability. Although a lot of progress has been made over the years, one critical challenge that the low-frequency UWB SAR technology still faces is the discrimination of targets of interest from other natural and manmade clutter objects in the scene. The key issue is that the targets of interests are typically small compared to the wavelengths of the radar signals in this frequency band and have very low radar cross sections (RCSs). Thus it is very difficult to discriminate targets and clutter objects using the low-resolution SAR imagery.

In this paper, we consider the problem of discriminating and classifying buried targets of interest (metal and plastic mines, 155-mm unexploded ordnance [UXO], etc.) from other natural and manmade clutter objects (soda can, rocks, etc.) in the presence of noisy responses from the rough ground surfaces for low-frequency UWB 2-D SAR images. For classification problems, sparse representation-based classification [4] (SRC) has been successfully demonstrated in other imagery domains such as medical image classification [5]–[8], hyperspectral image classification [9]–[11], high-resolution X-band synthetic aperture radar (SAR) image classification [12], recaptured image recognition [13], video anomaly detection [14], and several others [15]–[21]. However, in the low-frequency RF UWB SAR domain, although we have studied the feasibility of using SRC for higher-resolution 3-D down-looking SAR imagery [22], the application of SRC to low-frequency UWB 2-D SAR imagery has not been studied to date due to the aforementioned low-resolution issue. In this paper, we generalize the traditional SRC to incorporate the target classification using either a single channel (radar polarization) or multiple channels of SAR imagery.

In sparse representations, many signals can be expressed by a linear combination of few bases taken from a “dictionary”. Based on this theory, SRC [4] was originally developed for robust face recognition. The main idea in SRC is to represent a test sample (e.g., a face) as a linear combination of samples from the available training set. Sparsity manifests because most of the nonzero components correspond to bases whose memberships are the same as the test sample.

The aforementioned SRC is solely applied to the problem of one channel, using \(l_1\)-norm as the sparsity regularizer and without the presence of the shared class. Our main contribution: In this paper, we generalize the traditional SRC to a model that can use the information of the shared class, and implement multichannel classification problems by exploiting structures of sparse coefficients using various techniques. With the presence of the shared class, a new test sample is presumably represented by samples from the corresponding class in conjunction with samples from the shared class. Particularly, a noisy test sample of a target, e.g. T1, can be expressed as a linear combination of available clean training images from that target and some ground images at different
angles. This assumption still assures the sparsity of the coding coefficient. After obtaining the sparse coefficient, we can eliminate the contribution of the ground and focus on the discriminative information of the signal, which likely enhances the performance of the SRC framework. For multichannel signals, in order to use the cross-channel correlation, we present four different sparsity regularizers imposing on the sparse codes.

We also note here that our proposed models are not limited at the problem of classifying UWB SAR images; they can be applied to any kind of signals which contain multiple channels, e.g., multiple color channels in medical image processing or multiple viewpoints of objects.

II. SPARSE REPRESENTATION-BASED CLASSIFICATION

A. Notation

Each target sample is represented by a SAR image formed using either a single (using co-pol) or multiple polarization (using both co-pol and cross-pol) channels. Thus, one target sample is denoted by \( y \in \mathbb{R}^{d \times 1 \times T} \), where \( d \) is the total number of image pixels and \( T \) is the number of polarization channels. It is noted that \( y \) may represent a tensor when \( T > 1 \). A collection of \( N \) samples is denoted by \( Y \in \mathbb{R}^{d \times N \times T} \).

Suppose we are considering a general classification problem with \( C \) different classes, let \( D_c (1 \leq c \leq C) \) be the collection of all training samples from class \( c \), \( D_0 \) be the collection of samples in the shared class, and \( D = [D_1, \ldots, D_C, D_0] \) be the total dictionary. In our problem, the shared class can be seen as the collection of ground images.

For any tensor \( M \), let \( M^{(t)} \) be its \( t \)-th channel. For convenience, given two tensors \( M, N \), the tensor multiplication \( P = MN \) is considered as channel-wise multiplication, i.e., \( P^{(t)} = M^{(t)}N^{(t)} \). For a tensor \( M \), we also denote the sum of square of all elements by \( \|M\|_F \) and the sum of absolute values of all elements by \( \|M\|_1 \). Tensor addition/subtraction simply represents element-wise addition/subtraction.

With the definition of tensor multiplication, a sparse representation of \( y \) using \( D \) can be obtain by solving:

\[
x = \arg \min_x \frac{1}{2}\|y - Dx\|_F^2 + \lambda \mathcal{F}_S(x)
\]

where \( \lambda \) is a positive regularization parameter and \( \mathcal{F}_S(x) \) is a function that encourages a sparsity structure of \( x \). Denote by \( x^t \) the sparse coefficient of \( y \) on \( D_t \). Then, the tensor \( x \) can be divided into \( C + 1 \) tensor parts \( x^1, x^2, \ldots, x^C, x^0 \).

B. Generalized sparse representation-based classification

Confuser detection:

In practical problems, the set of confusers is not limited in the training set. A confuser can be anything that is not a target; it can be solely the ground or a capture of an unseen object. In the former case, the test signal can be well presented by using only the ground \( D_0 \), while in the latter case, the sparsity assumption is no longer valid. Therefore, one test signal \( y \) is classified as a confuser if one of the following three conditions satisfies: i) it is not sparsely interpreted by the total dictionary \( D \); ii) it has the most active elements in the sparse code locating at \( x^0 \); and iii) it is similar to known confusers.

![Fig. 1: Different sparsity constraints on x.](image)

Algorithm 1 Generalized sparse representation-based classification with a shared class

```
function IDENTITY(y) =
    GENERALIZED_SRC(y, D, λ, \mathcal{F}_S(\bullet), \varepsilon, \tau)

INPUT:
y \in \mathbb{R}^{d \times 1 \times T} - a test signal;
D = [D_1, D_2, \ldots, D_C, D_0] \in \mathbb{R}^{d \times K \times T} - the total dictionary with the shared dictionary \( D_0 \);
\mathcal{F}_S(\bullet) - the sparsity constraint imposed on sparse codes.
λ \in \mathbb{R}^+ - a positive regularization parameter;
\varepsilon, \tau - positive thresholds.

OUTPUT: 
the identity of \( y \).
1. Sparsely code \( y \) on \( D \) via solving:
   \[
x = \arg \min_x \{\|y - Dx\| + \lambda \mathcal{F}_S(x)\}
   \]
2. Remove contribution of the shared dictionary:
   \( \bar{y} = y - D_0 x^0 \).
3. Calculate class-specific residuals:
   \[r_c = \|y - D_c x^c\|_F, \forall c = 1, 2, \ldots, C.\]
4. Decision:
   if \( \min_c(r_c) > \tau \) (an unseen object) or \( \|\bar{y}\|_2 < \varepsilon \) (a ground) then
     \( y \) is a confuser.
   else
     \( y \) is a target.
```

The overall algorithm of general SRC applied to multichannel signals with presence of a shared class is shown in Algorithm 1.

The overall algorithm of generalized SRC applied to multichannel signals in the presence of a shared class is shown in Algorithm 1. In this algorithm, \( \mathcal{F}_S(x) \) is a sparsity regularizer imposed on \( x \), which can be defined in different ways:

a) Apply a sparsity constraint on each column of the coefficient tensor \( x \), no cross-channel constraint.

In this case, active elements of \( x \) could be randomly distributed. We can see that it is similar to solving \( T \) separate sparse coding problems, each for a polarization channel. The
classification rule is executed based on the sum of all the squares of the residuals. We refer to this framework as SRC-cumulative residual or SRC-CR (CR in the short version). (See Fig. 1a with sparse tensor $x_1$). b) Concatenate all channels.

The most convenient way to convert a multichannel to a single-channel problem is to concatenate all $T$ channels of one signal to obtain a long vector. By doing so, we have the original SRC framework with $l_1$-norm minimization. Consequently, sparse codes of all $T$ channels are identical. From the tensor viewpoint, the code tensor $x$ will have few active “tubes”; moreover, all elements in a tube are the same. We refer to this framework as SRC-concatenation or SRC-CC (CC in the short version). (See Fig. 1b with sparse tensor $x_2$).

c) Use a simultaneous model.

Similar to the SHIRC model proposed in [6], [8] we can impose one constraint on active elements of tensor $x$ as follows: $x$ also has few nonzero tubes as in SRC-CC; however, elements in one active tube are not necessarily the same. In other words, the locations of nonzero coefficients of training samples in the linear combination exhibit a one-to-one correspondence across channels. If the $j$-th training sample in $D^{(1)}$ has a nonzero contribution to $y^{(1)}$, then for $t \in \{2, \ldots, T\}$, $y^{(t)}$ has a nonzero contribution from the $j$-th training sample in $D^{(t)}$. We refer to this framework as SRC-Simultaneous or SRC-SM (SM in the short version). (See Fig. 1c with sparse tensor $x_3$). To achieve this requirement, we can impose on the tensor $x$ (with one column and $T$ channels) the $l_{1,2}$-minimization constraint, which is similar to the row-sparsity constraint applied on matrices in [12].

Remarks: While SHIRC uses $l_0$-minimization on $x$ and applies the modified SOMP [23], our proposed SRC-SM exploits the flexibility of $l_{1,2}$-regularizer since it is easily modified when more constraints present (e.g., non-negativity). In addition, it is more efficient especially when dealing with problems of multiple samples at input.

d) Use a Group tensor model.

Intuitively, since one object is ideally presented by a linear combination of the corresponding dictionary and the shared dictionary, it is highly expected that number of active (tensor) parts in $x$ is small, i.e., only few of $x^1, \ldots, x^5$ are nonzero tensors. This suggests us a group tensor sparsity framework that can improve the classification performance, which is referred to as SRC-GT (GT in the short version). The visualization of this idea is shown in Fig. 1d.

C. Problem formulation and solution

While solving the main optimization problems in SRC-CC and SRC-CR are pretty straightforward, those in SRC-SM and SRC-GT require more explanation.

In order to achieve the “tube-sparsity” (SRC-SM) and “group-tensor sparsity” (SRC-GT) solutions, we propose the following optimization problem:

$$x = \arg \min_{x} \frac{1}{2} \| y - Dx \|_2^2 + \lambda \sum_{g_i \in \mathcal{G}} \| x_{g_i} \|_2$$

(3)

where $\mathcal{G}$ is a partition of the tensor $x$, and $x_{g_i}$ is the vectorization of the part corresponding to $g_i$. In SRC-SM, $\mathcal{G}$ is divided into tubes, while in SRC-GT, $\mathcal{G}$ is partitioned into all tubes in the corresponding class, e.g., $x_{g_i} = x^i$.

This problem is similar to the group lasso problem proposed in [24]. The solution can be found using FISTA [25] and the minimum $l_2$-norm minimization, which will be discussed in details in our future works.

III. EXPERIMENTAL RESULTS

A. Data

The SRC is applied to a SAR database consisting of targets (metal and plastic mines, 155-mm unexploded ordinance [UXO], etc.) and clutter objects (soda can, rocks, etc.) buried under rough ground surfaces. The electromagnetic (EM) radar data are simulated based on the full-wave computational EM method known as finite-difference, time-domain (FDTD) software [26], which was developed by the U.S. Army Research Laboratory (ARL). The software was validated for a wide variety of radar signature calculation scenarios [27], [28]. Our volumetric rough ground surface grid – with the embedded buried targets – was generated by using the surface root-mean-square (rms) height and the correlation length parameters. The
targets are flush buried at 2-3 cm depth. In our experiments, the easiest case of rough ground surface in our experiments, the surface rms is 5.6 mm and the correlation length is 15 cm. The SAR images of various targets and clutter objects are generated from EM data by coherently integrated individual radar return signals along over a range of aspect angles. The SAR images are formed using the backprojection image formation [29] with an integration angle of 60°. Fig. 2a shows the SAR images (using VV polarization) of some targets that are buried under a perfectly smooth ground surface. Each target is imaged at a random viewing aspect angle and an integration angle of 60°. Fig. 2b and 2c shows the same targets Fig. 2a, except that they are buried under a rough ground surface (the easiest case corresponds to ground scale = 1 and harder case corresponds to ground scale = 5). Similarly, Fig. 2d, 2e and 2f show the SAR images of some clutter objects buried under a smooth and rough surface, respectively. For training, the target and clutter object are buried under a smooth surface in order to generate high signal-to-clutter ratio images. We include 12 SAR images that correspond to 12 the different aspect angles (0°, 30°, ..., 330°) for each target type. For testing, the SAR images of targets and confusers are generated at random aspect angles and buried under the rough ground surfaces. Various levels of ground surface roughness are simulated by selecting different ground surface scaling factors when embedding the test targets under the rough surfaces. Thus, the resulting test images are very noisy with a very low signal-to-clutter ratio. Each data sample of one object is represented by either i) one SAR image using data from one co-pol (VV, HH) channel or ii) two or more images using data from co-pol (VV, HH) and cross-pol (HV) channels. For each target type, we test 100 image samples measured at random aspect angles.

### B. Overall classification accuracy

We apply four presented methods to different possible combinations of three polarizations: VV, HH, VV-HH, VV-HV, HH-HV, and VV-HH-HV (or ALL), and also compare these results with those obtained by SVM using the libsvm library [30]. The training set comprises all five target sets, four out of five confuser sets (each time we leave one confuser set out, which is supposed to be unseen), and the ground set. Each target is considered as one class, all confusers are gathered into one class – class of confusers. The test set comprise of all test targets and confusers, each of them is contaminated by one of the test grounds at scale 1 – small disruption. We also impose the non-negative constraints on sparse coefficients to compare all results. The results are averaged and reported in Table I. From Table I, we can observe that i) the performance of each method is, in general, improved with non-negativity constraints imposed on sparse coefficients; ii) the HH polarization itself and all combinations containing it yield better results; and iii) when more than two polarizations are involved, SRC-SM outperforms all other methods. It is worth noting here that all SRC-like methods perform better than SVM does.

### C. Polarization combination comparison

To see which combination of polarizations provides good results, we conduct another experiment by applying each of competing methods on six polarization combinations. The data are set up as in section III-B with slightly modifications: i) we use the non-negativity constraints, which were observed to have better performance; and ii) in test signals, the ground scale is varied from 1 to 5 to see how each method performs with high corruptions. Fig. 3 shows the results of all methods.
on different combinations. First, among all methods, VV-HV, HH-HV, and ALL combinations outperform others with the gaps being significant when ground scales increase. It is also evident that HH-HV consistently outperforms other combinations in all methods. Second, among all four presented methods, the SRC-CC results are clearly bettered by others with its best number being under 90%. Third, SRC-GT using HH only also provides good results; it is even the best when the ground scale is equal to 1. SVM again does not provide good performances.

D. Classification results as a function of training confuser sets

In this experiment, we vary the number of training confuser sets from one to four in conjunction with different ground scales. The results are shown in Fig. 4. It is evident that with small corruption, SRC-based methods provide almost the same results with GT-HH being slightly better than others. However, when the ground scale is increased, SM-HHHV and SM-ALL offer the best results.

IV. CONCLUSION

We have applied four sparse representation-based classifiers to identify and discriminate buried targets of interest from other natural and manmade clutter objects in the presence of noisy responses from the rough ground surfaces for low-frequency UWB 2-D SAR images. The application of using SRC for low-frequency UWB 2-D SAR imagery has never been studied. The discrimination/classification results are very encouraging, even when the test images are very noisy (buried under rough ground surfaces), targets have small RCS, and the target’s responses have very little detailed structures (targets are small compared to the wavelengths of the radar signals). The simultaneous sparse-based (SM) technique consistently offers the best results with the combination of co- and cross-pol data (e.g., HH-HV). On the other hand, the group tensor sparse-based (GT) technique outperforms others when only one polarization (HH) is available. In addition, the non-negativity constraints on sparse codes enhance the performance of the system. In future works, we aim to incorporate dictionary learning frameworks into these models to obtain performance improvement.

REFERENCES


