Sparse representation of GPR traces with application to signal classification

Wenbin Shao  
*University of Wollongong, wenbin@uow.edu.au*

Abdesselam Bouzerdoum  
*University of Wollongong, bouzer@uow.edu.au*

Son Lam Phung  
*University of Wollongong, phung@uow.edu.au*

Publication Details  
Sparse representation of GPR traces with application to signal classification

Abstract
Sparse representation (SR) models a signal with a small number of elementary waves using an overcomplete dictionary. It has been employed for a wide range of signal and image processing applications, including denoising, deblurring, and compression. In this paper, we present an adaptive SR method for modeling and classifying ground penetrating radar (GPR) signals. The proposed method decomposes each GPR trace into elementary waves using an adaptive Gabor dictionary. The sparse decomposition is used to extract salient features for SR and classification of GPR signals. Experimental results on real-world data show that the proposed sparse decomposition achieves efficient signal representation and yields discriminative features for pattern classification. © 1980-2012 IEEE.

Keywords
classification, signal, application, representation, traces, sparse, gpr

Disciplines
Engineering | Science and Technology Studies

Publication Details

This journal article is available at Research Online: http://ro.uow.edu.au/eispapers/1160
Sparse representation models a complex signal with a small number of elementary waves using an over-complete dictionary. It has been employed for a wide range of signal and image processing applications. In this paper, we present an adaptive sparse representation method for modeling and classification of ground penetrating radar (GPR) signals. The proposed method decomposes each GPR trace into elementary waves using a dynamically expanding Gabor dictionary. The sparse decomposition is used to extract salient features for classification of GPR signals. Experimental results on real-world data show that the proposed sparse decomposition achieves efficient signal representation, and yields discriminative features for pattern classification.

Index Terms

Sparse representation, signal decomposition, ground penetrating radar, pattern classification.
I. INTRODUCTION

Ground penetrating radar (GPR) is often used for non-destructive geophysical testing. It probes the subsurface area with electromagnetic waves. The characteristics of underground objects are identified through pseudo-imaging and signal processing. GPR has become a valuable tool in several applications, such as archaeological explorations [1], glacier and ice sheet investigation [2, 3], detection and monitoring of below-ground biological structures [3], mineral exploration and resource evaluation [4], building condition assessment [2], road pavement analysis [3, 5], and landmine detection [6].

This paper addresses the problem of sparse representation (SR) of GPR signals. Sparse representation aims to find an efficient signal decomposition by expressing a signal as a linear combination of few signal atoms chosen from an over-complete dictionary. A related area to sparse representation is compressed sensing (CS) theory, which affirms that sparse signals can be reconstructed from under-sampled information [7, 8]. Both SR and CS have been employed in numerous signal and image processing applications, such as denoising [9], deblurring [10], compression [11], and reconstruction [12]. For example, sparse representation was used in hyperspectral imagery for modeling, source separation, mapping, and classification [13, 14]. Tang et al. applied SR to wideband beamforming for direction-of-arrival (DOA) estimation [15]; they were able to extract the target DOAs without ambiguity. In aerospace remote sensing, compressed sensing was employed to deblur highly incomplete measurements [16]. For a more comprehensive treatment of CS theory and applications, the reader is referred to [17, 18].

In radar applications, CS theory has been applied to radar imaging [19, 20, 21, 22], radar signal processing [23], and radar design [18, 24]. Gurbuz et al. presented a CS-based data acquisition and imaging approach for ground penetrating radar [25], and later they extended the CS imaging approach to stepped-frequency GPRs [26]. Yoon and Amin applied CS to through-the-wall radar imaging (TWRI) [27]. Qu and Yang proposed a CS migration imaging method for SFCW GPR system to address the issues of strong air/ground interface reflection and finite antenna beamwidth [28]. Soldovieri et al. proposed a sparse minimization algorithm for GPR rebar detection [29]. Yang et al. proposed an approach for multi-view TWRI using compressed sensing [30]. Their experimental results show that their approach, which combines image formation and fusion, achieves better reconstruction accuracy compared to the approach of image formation followed
Suksmono et al. applied CS theory to the selection of frequency measurements for an SFCW GPR systems [31]; they found that the CS based design can acquire data eight times faster than the traditional SFCW GPR. Huang et al. presented a data acquisition scheme and an imaging algorithm for ultra-wideband TWRI based on compressed sensing [32].

In sparse representation, the choice of the dictionary plays a crucial role in the signal decomposition. Approaches for dictionary construction in sparse representation fall into two main categories: model-based and learning-based [33]. McClure and Carin proposed a matching pursuits method using a wave-based dictionary for scattering data [34]. The dictionary comprises atoms of wavefronts, resonances, and chirps. Their results show rapid convergence even in the presence of high noise. However, their approach requires knowledge of the incident-pulse shape, the resonant frequencies and chirp frequencies in advance. In this paper, we present an adaptive dictionary construction approach for GPR signal representation, where the resonance frequencies are unknown. In a GPR survey, particular resonance frequencies arise in wave propagation; therefore, reflected waves from different buried objects or paths present different electromagnetic characteristics. Furthermore, GPR signals approximately resemble the Ricker wave (second-order derivative of Gaussian) [1, 35, 36]. Inspired by these observations, we propose to represent the GPR signals using an adaptive Gabor dictionary. Preliminary results of the proposed adaptive signals decomposition and its application to classification of railway ballast traces were presented in [37]. In this paper, we improve the decomposition procedure, enrich the feature extraction approach, and present more comprehensive experimental results. The proposed signal decomposition method is also compared to the wavelet decomposition and K-SVD, a learning dictionary method [38].

The paper is organized as follows. Section II describes the GPR system and the data sets used in the experimental methods. Section III gives a brief introduction to sparse signal representation, describes the proposed signal decomposition method, and analyzes its effectiveness in GPR signal representation. Section IV addresses the problem of GPR signal classification using feature extraction based on the proposed signal decomposition. Section V gives the concluding remarks.

II. GPR SYSTEM OVERVIEW AND EXPERIMENTAL DATA SETS

This section gives a brief overview of a GPR system, the GPR signals, and the data preprocessing stage. It also introduces the GPR data sets used in the experimental evaluation.
A. GPR system

A GPR system consists of a transmitter (signal generator), transmitting and receiving antennas, and a receiver (recording device) [2, 39]. To detect underground objects using GPR, the transmitter generates an electromagnetic pulse. The electromagnetic wave radiates from the transmitting antenna into the subsurface. If there is an object on the path of the wave propagation whose electrical properties are different from those of surrounding materials, part of the wave energy is reflected back. The reflected energy is detected by the receiving antenna and processed by the receiver. The receiver starts recording after a pulse has left the transmitting antenna and stops after a certain time window has elapsed. The recorded pulse sequence as a function of time is called a trace. Successive traces displayed side by side form a pseudo-image, known as B-scan (or time-distance record, space-time data). Figure 1 shows a GPR B-scan and a trace.

![GPR profile B-scan display. The vertical line on the left indicates where the trace on the right is obtained. See the electronic edition for a color version of this figure.](image)

In this paper all GPR data were pre-processed using several techniques, including DC component removal, re-sampling and time shifting. DC component removal subtracts the mean of each trace to reduce the intrinsic interference of the system; re-sampling is to ensure sampling rate consistency of the time-domain signals; time shifting aligns the signal based on the peak location of each trace.

B. Experimental data sets

Experiments in this paper were conducted on two GPR data sets: Windmill Islands data set and Wollongong railway data set. The Windmill Islands data set was collected from the Antarctic rocky islands [40]. It comprises GPR signals from three different surveys: Old Casey road GPR
survey, Loken Moraine GPR survey, and Wilkes GPR survey. The Old Casey road survey was aimed at imaging the bedrock height and examining road materials placed in previous years. The Loken Moraine GPR survey was conducted to probe the structures related to moraines development. The Wilkes GPR survey targeted cultural features for waste management. Various GPRs were used in the surveys with different antenna frequencies. In our experiments, we used only a subset of the data containing 300 samples acquired with an antenna frequency of 250 MHz.

The Wollongong railway data set was collected in our project for railway ballast assessment [41]. The aim of the project was to develop an automatic and non-destructive method using GPR for evaluating the conditions of railway ballast. The GPR surveys were conducted along an existing railway track in Wollongong, New South Wales, Australia. The experimental track used was parallel to several tracks that were in service. Considering the time and cost, three railtrack sections with known ground truth were used for ballast condition assessment. Each section had a length of 2.0 m and a depth of 0.55 m; the width was equivalent to the existing ballast width. To target the most common ballast fouling conditions, three ballast types were considered: clean, 50% clay fouling, and 50% coal fouling. Here, the fouling material was measured using relative ballast fouling ratio. The radar antenna frequency used was 800 MHz.

The entire Wollongong data set with known ground truth has three subsets based on the antenna heights $h$. The antenna heights for Set 1, Set 2, and Set 3 were 200 mm, 300 mm, and 400 mm, respectively. Set 1 and Set 2 were collected under dry ground conditions: sunny weather and dry materials. Set 3 was acquired under wet conditions: cloudy weather and water-saturated materials. A summary of the Wollongong railway data set is presented in Table I.

**TABLE I**

**Settings and numbers of available traces in the Wollongong railway data set.**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Settings</td>
<td>$h = 200$ mm, dry</td>
<td>$h = 300$ mm, dry</td>
<td>$h = 400$ mm, wet</td>
</tr>
<tr>
<td>Section clean (traces)</td>
<td>477</td>
<td>478</td>
<td>642</td>
</tr>
<tr>
<td>Section clay (traces)</td>
<td>469</td>
<td>470</td>
<td>745</td>
</tr>
<tr>
<td>Section coal (traces)</td>
<td>436</td>
<td>438</td>
<td>705</td>
</tr>
<tr>
<td>Total traces</td>
<td>1382</td>
<td>1386</td>
<td>2092</td>
</tr>
</tbody>
</table>
III. ADAPTIVE SPARSE DECOMPOSITION AND ANALYSIS

In this section, we present the proposed adaptive sparse signal decomposition, and evaluate its effectiveness in GPR signal representation. Before introducing the proposed adaptive GPR signal decomposition, we first present a brief review of sparse signal representation.

A. Sparse signal representation

Sparse representation expresses a signal as a linear combination of elementary waves. The elementary waves, called atoms, are chosen from an over-complete dictionary \( \mathbf{D} \in \mathbb{R}^{N \times M} \), with \( N < M \). The sparsity of a discrete-time signal \( \mathbf{x} \in \mathbb{R}^M \) is defined as the number of non-zero elements in \( \mathbf{x} \). The \( \ell_0 \) pseudo-norm, denoted as \( \| \mathbf{x} \|_0 \), is usually used as a measure of sparsity. If \( \| \mathbf{x} \|_0 = k \), the vector \( \mathbf{x} \) is called \( k \)-sparse. Suppose that the signal \( \mathbf{s} \) is to be modeled with a smaller number of atoms from the dictionary \( \mathbf{D} \). This can be formulated as a sparse representation problem

\[
P_0 : \min_{\mathbf{x}} \| \mathbf{x} \|_0 \text{ subject to } \mathbf{s} = \mathbf{Dx}.
\]  

The combinatorial optimization problem \( P_0 \) of finding a sparse solution has been shown to be NP-hard [18, 42]. Unlike the \( \ell_2 \)-norm optimization, we cannot solve problem \( P_0 \) directly using convex analysis because the \( \ell_0 \) “norm” is discrete and discontinuous. Therefore, in practice two types of algorithms are usually used: greedy methods and convex relaxation algorithms. The greedy methods iteratively approximate the signal. At each iteration, one atom is chosen that maximally reduces the \( \ell_2 \) norm of the residual error. The two most widely used algorithms in this category are matching pursuit [43] and orthogonal matching pursuit [44, 45].

Convex relaxation methods replace the \( \ell_0 \) “norm” by a related convex approximation. Basis pursuit is the main technique for convex relaxation [18, 46]; it relaxes the \( \ell_0 \) “norm” using the \( \ell_1 \) norm. Therefore, problem \( P_0 \) becomes

\[
P_1 : \min_{\mathbf{x}} \| \mathbf{x} \|_1 \text{ subject to } \mathbf{s} = \mathbf{Dx}.
\]  

The convex optimization problem \( P_1 \) can be solved by several software tools, such as \( \ell_1 \)-magic [47] and CVX [48].
B. Adaptive sparse decomposition of GPR signals

In the proposed sparse decomposition, a GPR trace is decomposed into delayed and scaled Gabor wavelets. That is, a radar trace \( s(t) \) is expressed as a linear combination of elementary waves

\[
s(t) = \sum_{i=1}^{K} \alpha_i g_i(t - \tau_i) + \nu(t)
\]  

(3)

where \( \alpha_i \) is a scalar weight, \( g_i(t - \tau_i) \) is a Gabor atom with a time delay \( \tau_i \), and \( \nu(t) \) is a residual signal that we aim to minimize. There are two types of Gabor atoms, even and odd Gabor functions:

\[
g_i(t) = \begin{cases} e^{-\frac{t^2}{2\sigma_i^2}} \cos(2\pi f_it) : \text{even function,} \\ e^{-\frac{t^2}{2\sigma_i^2}} \sin(2\pi f_it) : \text{odd function,} \end{cases}
\]  

(4)

where \( \sigma_i \) is the standard deviation and \( f_i \) is the frequency.

In traditional sparse representation approaches, the dictionary \( D \) is constructed \textit{a priori}, then used to solve the sparse representation problem; however, there are also techniques, such as K-SVD, which learn the dictionary iteratively. In the proposed approach, the dictionary is based on Gabor wavelets, but it is not completely known \textit{a priori}. First, a Gabor dictionary is constructed using the atoms \( g_i(t) \) and used to perform an initial sparse signal decomposition; the delays \( \tau_i, (i = 1, \ldots, K) \), are considered unknown and must be determined adaptively for each selected atom. Furthermore, for each selected atom, the parameters, \( f_i \) and \( \sigma_i \), and the expansion coefficient \( \alpha_i \) are optimized using a search technique.

In the following, unless stated explicitly, all processing is performed in the discrete-time domain. Consider a GPR trace \( s \) consisting of \( N \) samples. The first step in the proposed adaptive decomposition is to build a dictionary of Gabor atoms, \( G = [g_1, g_2, \ldots, g_M] \) with all the functions \( g_i \) having unit norm and delays \( \tau_i = 0 \). The \( g_i \) parameters \( \sigma_i, f_i, \) and the length of atoms, are computed based on the GPR antenna frequency and sampling rate. This ensures that the dictionary is adaptive to the GPR signals.

The second step is to iteratively select the atom \( g_{i^*} \) that has maximum cross-correlation (in absolute value) with the residual signal:

\[
i^* = \arg \max_i \left[ \max_{\tau} |r_{ki}(\tau)| \right],
\]  

(5)
where \( r_{ki}(\tau) \) is the cross-correlation function between the residual signal \( \tilde{s}_{k-1} \) and the Gabor atom \( g_i(t) \). The optimum parameters of the selected atom \( g_i \) are then determined using a search technique by solving the following unconstrained optimization problem:

\[
\min_{\alpha*, \sigma*, f*} \frac{1}{2} \left[ \tilde{s}_{k-1} - \alpha* g_i(t - \tau*) \right]^2.
\]

(6)

Note that additional constraints can easily be incorporated into the atom selection process. For example, in addition to cross-correlation, energy ratio can be applied to search for an atom that fits a signal section first rather than the residual caused by imperfect fitting in previous iterations. To cope better with the computational complexity during implementation, we propose a hierarchical approach for finding the most appropriate atom at iteration. First, an atom is located using the correlation coefficients on the initial over-complete dictionary, see Eq.(5). Then, a new sub-dictionary is dynamically constructed based on the parameters of the selected atom \( g_i(t) \) and a second search is performed across the sub-dictionary. Next the atom with highest correlation coefficient is chosen and its delay is computed. The adaptive determination of time delay and hierarchical search ensures that the over-complete atom dictionary is dynamically expanding.

In the third step, the weights of all selected atoms are updated by solving

\[
\min_{\alpha_k} ||s - \Phi_k \alpha_k||_2,
\]

(7)

where \( \Phi_k \) consists of the time-delayed Gabor atoms that have been selected up to iteration \( k \). Finally, the residual signal is updated for the next iteration

\[
\tilde{s}_k = s - \Phi_k \alpha_k.
\]

(8)

This iterative procedure is repeated until a certain number of iterations is reached or the residual signal falls below a certain error tolerance level, \( \epsilon \),

\[
\frac{||\tilde{s}_k||_2}{||s||_2} < \epsilon.
\]

(9)

The detailed algorithm of the sparse signal decomposition is presented in Table II.

An example of the adaptive decomposition is shown in Fig. 2. The top plot shows the original GPR trace and its approximation (dashed line) using 15 atoms; the bottom plot shows the three atoms found in Iterations 1, 2, and 5. The atoms are illustrated with the computed time delays and the corresponding coefficients.
C. Analysis of sparse signal decomposition

In this subsection, we analyze the efficiency of the proposed sparse decomposition for GPR signal representation, and compare its performance to that of a discrete wavelet transform and the dictionary learning method K-SVD. Wavelet is a well-known approach for time-frequency decomposition. The k-SVD is a dictionary learning algorithm introduced by Aharon et al. [38]. Given a training set, the K-SVD iteratively updates the atoms in the dictionary to better fit the training data. It has been adopted in numerous applications.

In the sparse signal decomposition, both odd and even Gabor functions were used to build the initial dictionary. In the wavelet implementation, first the discrete wavelet transform with Daubechies wavelets of order 6 was applied to the GPR trace. The Daubechies wavelets were chosen because of their shape similarity with the GPR trace; in [49], Daubechies wavelets were used for feature extraction from GPR signals. Then, the wavelet coefficients were thresholded: only the coefficients larger than the threshold were kept, and the other coefficients were assigned to 0.

In the evaluation, 300 traces were randomly selected for comparison from each GPR data set. We calculated the normalized root mean squared error (NRMSE) for each trace. The NRMSE measure indicates the difference between the approximation signal and the original signal; it is
TABLE II

STEPS FOR SPARSE SIGNAL DECOMPOSITION.

1) Form a dictionary of atoms, \( G = [g_1, g_2, \cdots, g_M] \) with all the functions \( g_i \) having unit norm.
2) Initialize the iteration index \( k = 1 \), a residual signal \( \tilde{s}_0 = s \), and an empty matrix \( \Phi_0 = \emptyset \).
3) For the \( k \)-th iteration, compute the cross-correlation \( r_{ki}[\tau] \) of the function \( g_i \in G \) and the residual signal \( \tilde{s}_{k-1} \).
4) Find the atom \( g_{i^*} \) that gives the highest correlation, where
   \[
   i^* = \arg \max_i \left[ \max_{\tau} |r_{ki}(\tau)| \right],
   \tag{10}
   \]
   and determine the corresponding time delay \( \tau_k \).
5) Calculate \( \varphi_k \) via \( \varphi_k = g_{i^*}[n-\tau_k](u[n] - u[n-N]) \), where \( u[n] \) is the unit step function.
6) Form the updated matrix \( \Phi_k \) by adding column \( \varphi_k \):
   \[
   \Phi_k = [\Phi_{k-1}, \varphi_k].
   \]
7) Compute the weight vector \( \alpha_k = [\alpha_1, \ldots, \alpha_k]^T \):
   \[
   \alpha_k = (\Phi_k^T \Phi_k)^{-1} \Phi_k^T s.
   \tag{11}
   \]
   Note that Eq. (11) updates all the weights calculated from the previous iteration.
8) Update the residual signal: \( \tilde{s}_k = s - \Phi_k \alpha_k \).
9) Repeat Steps 3 to 8 until \( k \) reaches a pre-defined limit, or the residual satisfies
   \[
   \frac{||\tilde{s}_k||_2}{||s||_2} < \epsilon,
   \tag{12}
   \]
   where \( \epsilon \) is a selected tolerance based on the signal-to-noise ratio (SNR). Given an SNR \( \gamma \), the tolerance is calculated as \( \epsilon = 10^{-\gamma/20} \).

defined as
\[
\text{NRMSE} = \frac{1}{\sigma_s} \sqrt{\sum_{i=1}^N (s_i - \hat{s}_i)^2 / N},
\tag{13}
\]
where \( s_i \) is the \( i \)-th element of the original signal \( s \), \( \hat{s} \) is the signal approximation, and \( \sigma_s \) is the standard deviation of \( s \).

Figure 3 presents the NRMSE as a function of the number of expansion coefficients of the sparse representation. The adaptive sparse decomposition has a more consistent performance than the discrete wavelet transform or K-SVD. The proposed method requires only six or eight expansion coefficients to reach an NRMSE of 0.10 on both data sets. Furthermore, it has the
lowest NRMSE on the Windmill Islands data set, see Fig. 3(a). By contrast, the discrete wavelet transform requires 15 coefficients to reach an NRMSE of 0.10 on the Windmill Islands data set (Fig. 3(a)), and it does not reach the same NRMSE level on the Wollongong data set (Fig. 3(b)). The K-SVD method achieves the lowest NRMSE on the Wollongong railway data set, but it has the worst performance on the Windmill Islands data set (Fig. 3(a)). Tables III and IV present the NRMSE values for eighteen expansion coefficients.

![Graphs showing NRMSE comparison](image)

(a) Windmill Islands data set  
(b) Wollongong railway data set  

Fig. 3. Overall NRMSE of sparse signal decomposition, discrete wavelet processing and K-SVD with OMP recovery on two data sets.

<table>
<thead>
<tr>
<th>Number of expansion coefficients</th>
<th>Sparse signal decomposition</th>
<th>Discrete wavelet transform</th>
<th>K-SVD with OMP recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.10</td>
<td>0.27</td>
<td>0.29</td>
</tr>
<tr>
<td>10</td>
<td>0.06</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>15</td>
<td>0.03</td>
<td>0.10</td>
<td>0.18</td>
</tr>
<tr>
<td>18</td>
<td>0.02</td>
<td>0.08</td>
<td>0.15</td>
</tr>
</tbody>
</table>

The experimental results indicate that the sparse decomposition represents the GPR signal more efficiently with fewer coefficients compared to the discrete wavelet transform. Compared
to the dictionary learning algorithm K-SVD, the proposed approach decomposes one trace into several individual elementary waves (Fig. 4(a)); this is beneficial to subsequent analysis. The parameters of the decomposition, such as delay, frequency and bandwidth of each Gabor fitting function can be retrieved from the sparse signal decomposition and employed for automatic pattern classification.

![Normalized amplitude](image1)
![Normalized amplitude](image2)

(a) Adaptive sparse decomposition  
(b) K-SVD

Fig. 4. The first three atoms found using the two methods: (a) the proposed adaptive sparse decomposition, and (b) OMP recovery using the K-SVD dictionary. The original GPR trace is shown in Fig. 2.
IV. GPR SIGNAL CLASSIFICATION

In this section, we present a GPR signal classification system based on the proposed sparse representation. The system is comprised of four major stages: pre-processing, sparse signal decomposition, feature extraction, and classification (see Fig. 5). Given a trace, the system extracts features through adaptive decomposition and sends the feature vector to a classifier.

![Block diagram of the proposed system for GPR signal classification.](image)

To assess the effectiveness of the proposed signal decomposition for GPR signal classification, we apply it to evaluate railway ballast conditions using Wollongong railway data set described in Section II-B. The aim was to classify GPR traces into different categories (clean, 50% clay fouled ballast, and 50% coal fouled ballast) based on the ballast conditions. For classification, each GPR trace is represented by a feature vector derived from the adaptive signal decomposition. The feature vector is then used as input to a classifier. To evaluate the classifier generalization ability, we use five-fold cross validation. In the next subsection, we explain the feature extraction process. In Subsection IV-B, we present the results of GPR signal classification using the extracted features. In Subsection IV-C we compare the classification performance using the sparse decomposition with other feature extraction methods.

A. Feature extraction

A GPR signal captures the electromagnetic characteristics of reflectors (underground objects). The same information is contained in the parameters of the Gabor atoms that are chosen to represent the GPR signal. It is therefore logical to classify the GPR traces based on the parameters extracted from the sparse representation.

Based on the parameters derived from the adaptive signal decomposition, we propose a compact feature vector derived from decomposition for classification. Each trace is decomposed into a number of atoms; a feature vector consisting of atom parameters is extracted. The GPR data
are pre-processed so that traces have the same number of samples and the same sampling rate. Consider a GPR trace $s$, and let $K$ be the number of iterations in the sparse signal decomposition. From each Gabor atom, we extract its time delay $\tau_i$ ($i = 1, 2, \ldots, K$), frequency $f_i$, the width parameter $\sigma_i$, and the square of the expansion coefficient $\alpha_i$. Therefore, we have four sets of parameters that can be used for classification: the time delays $\{\tau_1, \tau_2, \ldots, \tau_K\}$, the frequencies of the Gabor atoms $\{f_1, f_2, \ldots, f_K\}$, the Gaussian width parameters $\{\sigma_1, \sigma_2, \ldots, \sigma_K\}$, and the expansion weights $\{\alpha^2_1, \alpha^2_2, \ldots, \alpha^2_K\}$. Since the first iteration always extracts the wave reflected from the surface of the ground, it is not used in the classification. The frequency feature vector $f$ is obtained by dividing the atom frequencies by the antenna frequency $f_a$.

$$f = [f_2, f_3, \ldots, f_K]^T / f_a.$$  \hfill (14)

The $\sigma$ feature vector is obtained by normalizing $\sigma_i$ by their mean value $\sigma_m$.

$$\sigma = [\sigma_2, \sigma_3, \ldots, \sigma_K]^T / \sigma_m.$$  \hfill (15)

The energy feature vector $\alpha^2$ is obtained by dividing the order of magnitude of $\alpha_i^2$ values $O_m$.

$$\alpha^2 = [\alpha^2_2, \alpha^2_3, \ldots, \alpha^2_K]^T / O_m.$$  \hfill (16)

The delay feature vector $\tau$ is obtained by subtracting the first delay element from each subsequent delay,

$$\tau = [\tau_2 - \tau_1, \tau_3 - \tau_1, \ldots, \tau_K - \tau_1]^T.$$  \hfill (17)

The delay vector is also normalized by dividing it by 100 scale so that the value of the feature value corresponding to the largest delay lies close to one. At this stage, each element in every feature set is in the order of the decomposition index. To be used for classification, all feature vectors are sorted in descending order of the values of $\alpha^2$.

\begin{subsection}{Classification analysis}

The final stage of the proposed system is classification. There are many pattern classifiers, such as linear discriminant analysis, $k$-nearest neighbor ($k$-NN), Bayes classifier, neural networks, and support vector machines (SVMs) [50], which could be used to classify the extracted feature vectors. In this paper, we choose support vector machines as the classification tool because of their excellent generalization performance in various practical applications [51, 52, 53]. SVMs
are originally formulated for two-class problems. To handle multi-class classification, we use pair-wise SVMs.

The classification rates obtained using five-fold cross validation for single feature vectors are shown in Fig. 6. Different numbers of coefficients are evaluated for each feature set. Note that there are three classes correspond to three ballast types: clean (Class 1), 50% clay fouling (Class 2), and 50% coal fouling (Class 3). The classification rate is the ratio of the number of correctly classified traces to the total number of traces in the data set.

![Graphs showing classification performance](image)

Fig. 6. Classification performance on the three data sets using single feature set.

On Set 1, the feature vectors \( \tau \) and \( \alpha^2 \) have a better overall performance than the feature vectors
f and σ. When only 4 coefficients are used, τ, α² and f are able to achieve a classification rate above 80.0%. On Set 2, the feature vector τ performs the best; it has a classification rate of 87.2% with only 4 coefficients. When 4 to 8 coefficients are used, the feature vectors f and α² achieve similar performance. On Set 3, all feature vectors give good classification rates. The delay feature vector τ outperforms the others when few coefficients are used; it yields a classification rate of 94.8% with only 3 coefficients. Overall, the feature vectors τ, f, α² have an overall better performance than σ on the three data sets. Moreover, all feature vectors perform better on Set 3. On all sets, the classification performance on an individual class is close to the overall rate. The experimental results show that the parameters derived from the adaptive sparse decomposition are effective in classification.

The classification performance can be improved by combining different feature vectors. Our experiments show that, the combination of all four feature sets (τ, f, α², σ) achieves the best overall performance. Table V shows the classification rates for each ballast class and the overall classification rates using the combined feature vector on Set 1, Set 2, and Set 3, respectively. In the tables, the number of coefficients indicates the number of elements in one feature set. Because the combined vector consists of four feature sets, the total number of coefficients used for classification is four times the number given in the table. The 95% confidence interval using the Student’s t distribution is also reported when calculating the overall classification rate [54].

<table>
<thead>
<tr>
<th>Number of coefficients</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>94.2±1.5</td>
<td>91.7±1.8</td>
<td>99.3±0.5</td>
</tr>
<tr>
<td>5</td>
<td>93.8±1.6</td>
<td>91.2±1.8</td>
<td>97.6±0.8</td>
</tr>
<tr>
<td>7</td>
<td>94.3±1.5</td>
<td>93.0±1.6</td>
<td>96.6±1.0</td>
</tr>
</tbody>
</table>

The results show that the classification performance is improved by combining different feature vectors. With only 3 coefficients from each feature set, the classification rates are 94.2%, 91.7%, and 99.3% on Set 1, Set 2, and Set 3, respectively. The combined feature vector is also evaluated on the combination of Set 1, Set 2, and Set 3. The classification rates are 94.5% with 3 coefficients from each feature set and 93.6% with 5 coefficients from each feature set. The experimental results show that 3 coefficients from each feature set is sufficient for feature combination.
C. Comparison with other features

In this section, we compare the classification performance of the sparse decomposition features with wavelet and short-time Fourier transform (STFT) features. The sparse decomposition feature vector consists of the first three elements from each feature vector $\tau$, $f$, $\sigma$, and $\alpha^2$. For the wavelet features, the discrete-time wavelet transform is applied to each pre-processed GPR trace. The wavelet coefficients are then normalized by the mean absolute value, and the largest coefficients are selected to form the feature vector. The STFT features are extracted from the peaks of the spectrogram of the training data. The magnitudes of the peak spectra are normalized and arranged in descending order to form the feature vector. All three types of feature vectors are of the same length, i.e., they have 12 elements. The same pair-wise SVM configuration is used with all feature vectors, and five-fold cross validation is used to compute the overall classification rates. Table VI presents the classification rates for each data set. The sparse decomposition feature vector achieves the highest classification rate on two data sets, and it is very close to the STFT feature on Set 2. In summary, the sparse decomposition is very effective in signal classification.

<table>
<thead>
<tr>
<th>Features</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed combined feature vector</td>
<td>94.2±1.5</td>
<td>91.7±1.8</td>
<td>99.3±0.5</td>
</tr>
<tr>
<td>STFT</td>
<td>87.1±2.1</td>
<td>92.2±1.7</td>
<td>99.2±0.5</td>
</tr>
<tr>
<td>Wavelet</td>
<td>78.1±2.7</td>
<td>72.3±2.9</td>
<td>95.6±1.1</td>
</tr>
</tbody>
</table>

V. Conclusion

In this paper, we proposed an adaptive sparse decomposition for GPR signal analysis and classification. It employs an over-complete Gabor dictionary that is dynamically refined during the sparse decomposition. Furthermore, the proposed adaptive signal decomposition was found to be very effective for both signal representation and classification. Compared to the discrete wavelet transform and K-SVD, our sparse representation achieves better approximation of the GPR traces. The features extracted from the sparse signal decomposition were found to have a
high discrimination power in GPR signal classification; they outperform features extracted from wavelet decomposition and STFT.

ACKNOWLEDGMENT

This work is supported in part by a grant from the Australian Research Council. The Windmill Islands data were obtained from the Australian Antarctic Data Center (IDN Node AMD/AU), Australian Antarctic Division, Commonwealth of Australia. The Wollongong railway GPR data were collected as part of the Rail CRC-AT5 project, sponsored by CRC Rail for Innovation.

REFERENCES


