

Sparse Representation of GPR Traces With Application to Signal Classification

Wenbin Shao, Abdesselam Bouzerdoum, *Senior Member, IEEE*, and Son Lam Phung, *Member, IEEE*

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.

Index Terms—Ground penetrating radar (GPR), pattern classification, signal decomposition, sparse representation (SR).

I. INTRODUCTION

GROUND penetrating radar (GPR) is often used for nondestructive geophysical testing. It probes the subsurface area with electromagnetic waves. The characteristics of underground objects are identified through pseudoimaging 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. SR aims to find an efficient signal decomposition by expressing a signal as a linear combination of a few signal atoms chosen from an overcomplete dictionary. A related area to SR is compressed sensing (CS) theory, which affirms that sparse signals can be reconstructed from undersampled 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, SR 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, CS was employed to deblur highly incomplete measurements [16]. For a more comprehensive treatment of CS theory and applications, the reader is referred to [17] and [18].

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

In SR, the choice of the dictionary plays a crucial role in the signal decomposition. Approaches for dictionary construction in SR fall into two main categories: model based and learning based [33]. McClure and Carin proposed a matching pursuit 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 prior knowledge of the incident-pulse shape, the resonant frequencies, and chirp frequencies. 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 signal decomposition and its application to

Manuscript received September 10, 2012; accepted October 17, 2012. Date of publication February 1, 2013; date of current version June 20, 2013. This work was supported in part by a grant from the Australian Research Council.

The authors are with the Information and Communication Technology Research Institute and the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW 2522, Australia (e-mail: ws909@uowmail.edu.au; a.bouzerdoum@ieee.org; phung@uow.edu.au).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2012.2228660

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 singular value decomposition (K-SVD), a dictionary learning method [38].

This 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 the features extracted 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, on the path of the wave propagation, there is an object 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 pseudoimage, known as *B-scan* (or time–distance record, space–time data). Fig. 1 shows a GPR B-scan and a trace.

In this paper, all GPR data are preprocessed using several techniques, including dc component removal, resampling, and time shifting. DC component removal subtracts the mean of each trace to reduce the intrinsic interference of the system. Resampling is applied to ensure sampling rate consistency of the time-domain signals. Finally, 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

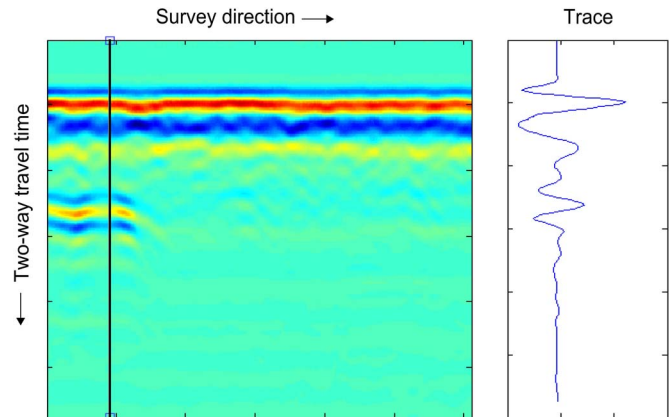


Fig. 1. GPR profile B-scan display. The vertical line on the left indicates where the trace on the right is obtained. See the electronic edition for a color version of this figure.

TABLE I
SETTINGS AND NUMBERS OF AVAILABLE TRACES
IN THE WOLLONGONG RAILWAY DATA SET

Data set	Set 1	Set 2	Set 3
Settings	$h = 200$ mm dry	$h = 300$ mm dry	$h = 400$ mm wet
Section clay	469	470	745
Section clean	477	478	642
Section coal	436	438	705
Total traces	1382	1386	2092

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. Our experiments used 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 nondestructive 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 analyze 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, 300, 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.

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

SR expresses a signal as a linear combination of elementary waves. The elementary waves, called atoms, are chosen from an overcomplete 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 nonzero elements in \mathbf{x} . The ℓ_0 pseudonorm, 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 small number of atoms from the dictionary \mathbf{D} . This can be formulated as an SR problem

$$P_0 : \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{s} = \mathbf{D}\mathbf{x}. \quad (1)$$

The combinatorial optimization problem P_0 of finding a sparse solution is nondeterministic polynomial time-hard [18], [42]. Unlike the ℓ_2 -norm optimization, we cannot solve problem P_0 directly using convex analysis because the ℓ_0 "norm" is discrete and discontinuous. Therefore, in practice, two types of algorithms are usually used: greedy algorithms and convex relaxation. The greedy algorithms iteratively approximate the signal. At each iteration, one atom is chosen that maximally reduces the ℓ_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 algorithms replace the ℓ_0 "norm" by a related convex approximation. Basis pursuit is the main technique for convex relaxation [18], [46]; it relaxes the ℓ_0 "norm" using the ℓ_1 norm. Therefore, problem P_0 becomes

$$P_1 : \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{s} = \mathbf{D}\mathbf{x}. \quad (2)$$

The convex optimization problem P_1 can be solved by several software tools, such as ℓ_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

$$s(t) = \sum_{i=1}^K \alpha_i g_i(t - \tau_i) + \nu(t) \quad (3)$$

where α_i is a scalar weight, $g_i(t - \tau_i)$ is a Gabor atom with a time delay τ_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

$$g_i(t) = \begin{cases} e^{-\frac{t^2}{2\sigma_i^2}} \cos(2\pi f_i t) & : \text{ even function} \\ e^{-\frac{t^2}{2\sigma_i^2}} \sin(2\pi f_i t) & : \text{ odd function} \end{cases} \quad (4)$$

where σ_i is the standard deviation and f_i is the frequency.

In traditional SR approaches, the dictionary \mathbf{D} is constructed *a priori* and then used to solve the SR 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 *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 τ_i ($i = 1, \dots, K$) are considered unknown and must be determined adaptively for each selected atom. Furthermore, for each selected atom, the parameters f_i and σ_i and the expansion coefficient α_i are optimized using a search technique.

In the following, unless stated otherwise, all processing is performed in the discrete-time domain. Consider a GPR trace \mathbf{s} consisting of N samples. The first step in the proposed adaptive decomposition is to build a dictionary of Gabor atoms, $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_M]$ with all the functions \mathbf{g}_i having unit norm and delays $\tau_i = 0$. The atom parameters σ_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 \mathbf{g}_{i^*} that has maximum cross-correlation (in absolute value) with the residual signal

$$i^* = \arg \max_i \left[\max_{\tau} |r_{ki}(\tau)| \right] \quad (5)$$

where $r_{ki}(\tau)$ is the cross-correlation function between the residual signal $\tilde{\mathbf{s}}_{k-1}$ and the Gabor atom $\mathbf{g}_i(t)$. The optimum parameters of the selected atom \mathbf{g}_{i^*} are then determined by solving the following unconstrained optimization problem:

$$\underset{\alpha_{i^*}, \sigma_{i^*}, f_{i^*}}{\text{minimize}} \frac{1}{2} [\tilde{\mathbf{s}}_{k-1} - \alpha_{i^*} \mathbf{g}_{i^*}(t - \tau_{i^*})]^2. \quad (6)$$

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 handle the computational complexity, during implementation, we propose a hierarchical approach for finding the most appropriate atom at each iteration. First, an atom is located using the correlation coefficients in an initial overcomplete dictionary, see (5). Then, a new subdictionary is dynamically constructed based on the parameters of the selected atom $\mathbf{g}_{i^*}(t)$, and a second search is performed across the subdictionary. Next, the atom with highest correlation coefficient is chosen, and its delay is computed.

TABLE II
 STEPS FOR SPARSE SIGNAL DECOMPOSITION

<p>1) Form a dictionary of atoms, $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_M]$ with all the functions \mathbf{g}_i having unit norm.</p> <p>2) Initialize the iteration index $k = 1$, a residual signal $\tilde{\mathbf{s}}_0 = \mathbf{s}$, and an empty matrix $\Phi_0 = \emptyset$.</p> <p>3) For the k-th iteration, compute the cross-correlation $r_{ki}[\tau]$ of the function $\mathbf{g}_i \in \mathbf{G}$ and the residual signal $\tilde{\mathbf{s}}_{k-1}$.</p> <p>4) Find the atom \mathbf{g}_{i^*} that gives the highest correlation, where</p> $i^* = \arg \max_i \left[\max_{\tau} r_{ki}(\tau) \right], \quad (10)$ <p>and determine the corresponding time delay τ_k.</p> <p>5) Calculate φ_k via $\varphi_k = \mathbf{g}_{i^*}[n - \tau_k](u[n] - u[n - N])$, where $u[n]$ is the unit step function.</p> <p>6) Form the updated matrix Φ_k by adding column φ_k:</p> $\Phi_k = [\Phi_{k-1}, \varphi_k].$ <p>7) Compute the weight vector $\alpha_k = [\alpha_1, \dots, \alpha_k]^T$:</p> $\alpha_k = (\Phi_k^T \Phi_k)^{-1} \Phi_k^T \tilde{\mathbf{s}}_k. \quad (11)$ <p>Note that Eq. (11) updates all the weights calculated from the previous iteration.</p> <p>8) Update the residual signal: $\tilde{\mathbf{s}}_k = \mathbf{s} - \Phi_k \alpha_k$.</p> <p>9) Repeat Steps 3 to 8 until k reaches a pre-defined limit, or the residual satisfies</p> $\frac{\ \tilde{\mathbf{s}}_k\ _2}{\ \mathbf{s}\ _2} < \epsilon, \quad (12)$ <p>where ϵ is a selected tolerance based on the signal-to-noise ratio (SNR). Given an SNR γ in dB, the tolerance is calculated as $\epsilon = 10^{-\gamma/20}$.</p>

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

$$\min_{\alpha_k} \|\mathbf{s} - \Phi_k \alpha_k\|_2 \quad (7)$$

where Φ_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

$$\tilde{\mathbf{s}}_k = \mathbf{s} - \Phi_k \alpha_k. \quad (8)$$

This iterative procedure is repeated until a selected number of iterations is reached or the residual signal falls below a predefined error tolerance ϵ

$$\frac{\|\tilde{\mathbf{s}}_k\|_2}{\|\mathbf{s}\|_2} < \epsilon. \quad (9)$$

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 figure shows the original GPR trace and its approximation (dashed line) using 15 atoms; the bottom figure shows the three atoms found in Iterations 1, 2, and 5. The atoms are shown 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

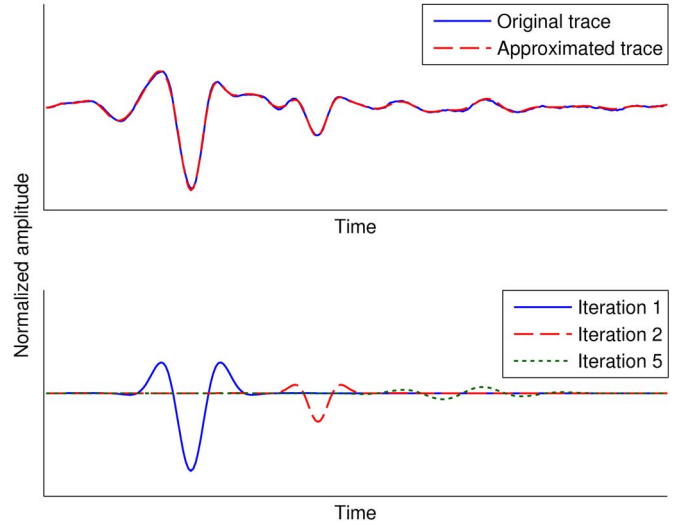


Fig. 2. Example of the proposed adaptive sparse decomposition: (Top plot) Original GPR trace and the approximated trace after $K = 15$ decomposition iterations and (bottom plot) the atoms found in Iterations 1, 2, and 5.

the K-SVD algorithm. The K-SVD, which has been adopted in numerous applications, 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.

In the proposed sparse signal decomposition, both odd and even Gabor functions were used to build the initial dictionary. In the wavelet decomposition, 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 also 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 zero.

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 defined as

$$\text{NRMSE} = \frac{1}{\sigma_s} \sqrt{\frac{\sum_{i=1}^N (s_i - \hat{s}_i)^2}{N}} \quad (13)$$

where s_i is the i th element of the original signal \mathbf{s} , $\hat{\mathbf{s}}$ is the signal approximation, and σ_s is the standard deviation of \mathbf{s} .

Fig. 3 shows the NRMSE as a function of the number of expansion coefficients of the SR. 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 for both data sets. Furthermore, it has the lowest NRMSE on the Windmill Islands data set. By contrast, the discrete wavelet transform requires 15 coefficients to reach an NRMSE of 0.10 on the Windmill Islands data set, and it does not reach the same NRMSE level on the Wollongong data set.

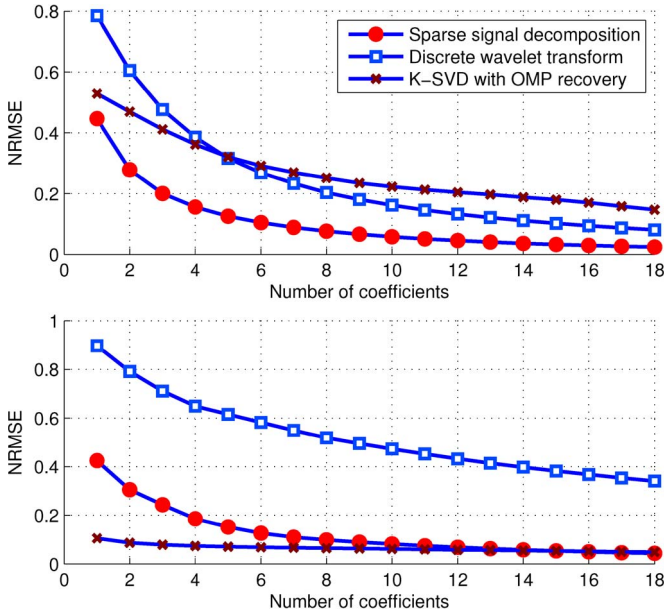


Fig. 3. Overall NRMSE of sparse signal decomposition, discrete wavelet processing, and K-SVD with orthogonal matching pursuit (OMP) recovery on (top plot) Windmill Islands data set and (bottom plot) Wollongong railway data set.

TABLE III
NRMSE OF SPARSE SIGNAL DECOMPOSITION, DISCRETE WAVELET PROCESSING, AND K-SVD WITH OMP RECOVERY ON WINDMILL ISLANDS DATA SET

Number of coefficients	Sparse signal decomposition	Discrete wavelet transform	K-SVD with OMP recovery
6	0.10	0.27	0.29
10	0.06	0.16	0.22
15	0.03	0.10	0.18
18	0.02	0.08	0.15

TABLE IV
OVERALL NRMSE OF SPARSE SIGNAL DECOMPOSITION, DISCRETE WAVELET PROCESSING, AND K-SVD WITH OMP RECOVERY ON WOLLONGONG RAILWAY DATA SET

Number of coefficients	Sparse signal decomposition	Discrete wavelet transform	K-SVD with OMP recovery
6	0.13	0.58	0.07
10	0.08	0.47	0.06
15	0.05	0.38	0.05
18	0.04	0.34	0.05

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. Tables III and IV present the NRMSE values as a function of the number of expansion coefficients.

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 pattern classification.

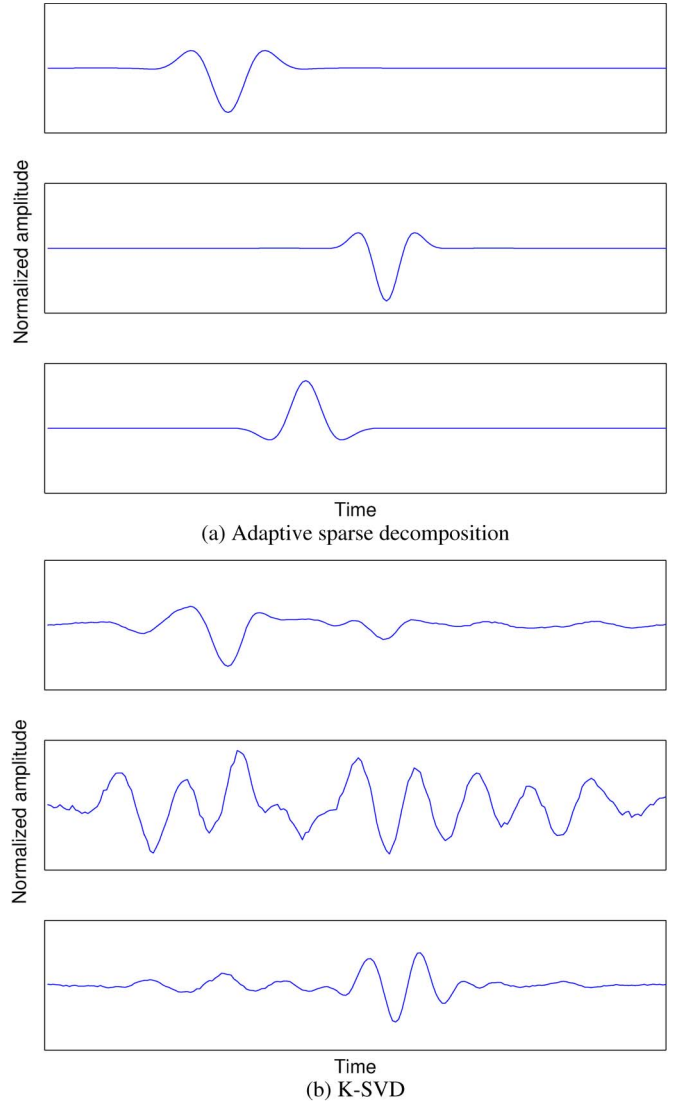


Fig. 4. First three atoms found using the two methods: (a) Proposed adaptive sparse decomposition and (b) OMP recovery using the K-SVD dictionary. The original GPR trace is shown in Fig. 2.



Fig. 5. Block diagram of the proposed system for GPR signal classification.

IV. GPR SIGNAL CLASSIFICATION

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

To evaluate the effectiveness of the proposed sparse decomposition for signal classification, we apply it to the classification of railway ballast conditions using Wollongong railway data set described in Section II-B. The aim is to classify GPR

traces into different categories (clean, 50% clay fouled ballast, and 50% coal fouled ballast) based on the ballast conditions. 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 fivefold cross-validation. In the next subsection, we explain the feature extraction process. In Section IV-B, we present the results of GPR signal classification using the extracted features. In Section IV-C, we compare the classification performance using the sparse decomposition with other feature extraction methods, namely, the wavelets and the short-time Fourier transform (STFT).

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 selected in the signal decomposition of GPR traces. It is therefore logical to classify the GPR traces based on the parameters extracted from the SR.

Before extracting the feature vector, the GPR data are pre-processed so that all traces have the same number of samples and the same sampling rate. Consider a GPR trace \mathbf{s} , and let K be the number of iterations in the sparse signal decomposition. From each Gabor atom, we extract its time delay τ_i ($i = 1, 2, \dots, K$), frequency f_i , the width parameter σ_i , and the square of the expansion coefficient α_i . Therefore, we have four sets of parameters that can be used for classification: the time delays $\{\tau_1, \tau_2, \dots, \tau_K\}$, the frequencies of the Gabor atoms $\{f_1, f_2, \dots, f_K\}$, the Gaussian width parameters $\{\sigma_1, \sigma_2, \dots, \sigma_K\}$, and the expansion weights $\{\alpha_1^2, \alpha_2^2, \dots, \alpha_K^2\}$. 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 \mathbf{f} is obtained by dividing the atom frequencies by the antenna frequency f_a

$$\mathbf{f} = \frac{1}{f_a} [f_2, f_3, \dots, f_K]^T. \quad (14)$$

The σ feature vector is obtained by dividing σ_i by the mean value of the Gaussian width parameters used in the Gabor dictionary σ_0

$$\sigma = \frac{1}{\sigma_0} [\sigma_2, \sigma_3, \dots, \sigma_K]^T. \quad (15)$$

The energy feature vector α^2 is obtained by dividing the coefficients α_i^2 by a constant α_0

$$\alpha^2 = \frac{1}{\alpha_0} [\alpha_2^2, \alpha_3^2, \dots, \alpha_K^2]^T. \quad (16)$$

The constant α_0 is chosen on the order of the square of the expansion coefficients $\alpha_0 = 10^8$. The delay feature vector τ is obtained by subtracting the first delay element from each subsequent delay

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

The delay vector is also normalized by dividing it by 100 so that the value of the feature corresponding to the largest delay is close to one. At this stage, each feature vector is ordered in terms of the decomposition index. However, the feature vectors used for classification are sorted in descending order of α^2 .

B. Classification Analysis

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

The classification rates obtained using fivefold 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 corresponding to three ballast types: clean (Class 1), 50% clay fouling (Class 2), and 50% coal fouling (Class 3). The classification rate is the percentage of traces in the data set that are correctly classified.

On Set 1, the feature vectors τ and α^2 have a better overall performance than the feature vectors \mathbf{f} and σ . When only four coefficients are used, feature vectors τ , α^2 , and \mathbf{f} are able to achieve a classification rate of above 80.0%. On Set 2, the feature vector τ performs the best; it has a classification rate of 87.2% with only four coefficients. When four to eight coefficients are used, the feature vectors \mathbf{f} and α^2 achieve similar performance. On Set 3, all feature vectors give good classification accuracy. The delay feature vector τ outperforms the others when only few coefficients are used; it yields a classification rate of 94.8% with only three coefficients. Overall, the feature vectors τ , \mathbf{f} , and α^2 have a better performance than σ on the three data sets. Moreover, all feature vectors perform better on Set 3. Furthermore, classification performance on individual classes is close to the overall classification accuracy. The experimental results show that the parameters derived from the adaptive sparse decomposition are effective for classification.

The classification performance can be improved by combining different feature vectors. Our experiments show that the combination of all four feature sets (τ , \mathbf{f} , α^2 , σ) achieves the best overall performance. Table V shows the overall classification rates using the composite 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 composite feature 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 for the overall classification rate [54].

With only three coefficients from each feature set, the classification rates reach 94.2%, 91.7%, and 99.3% on Set 1, Set 2,

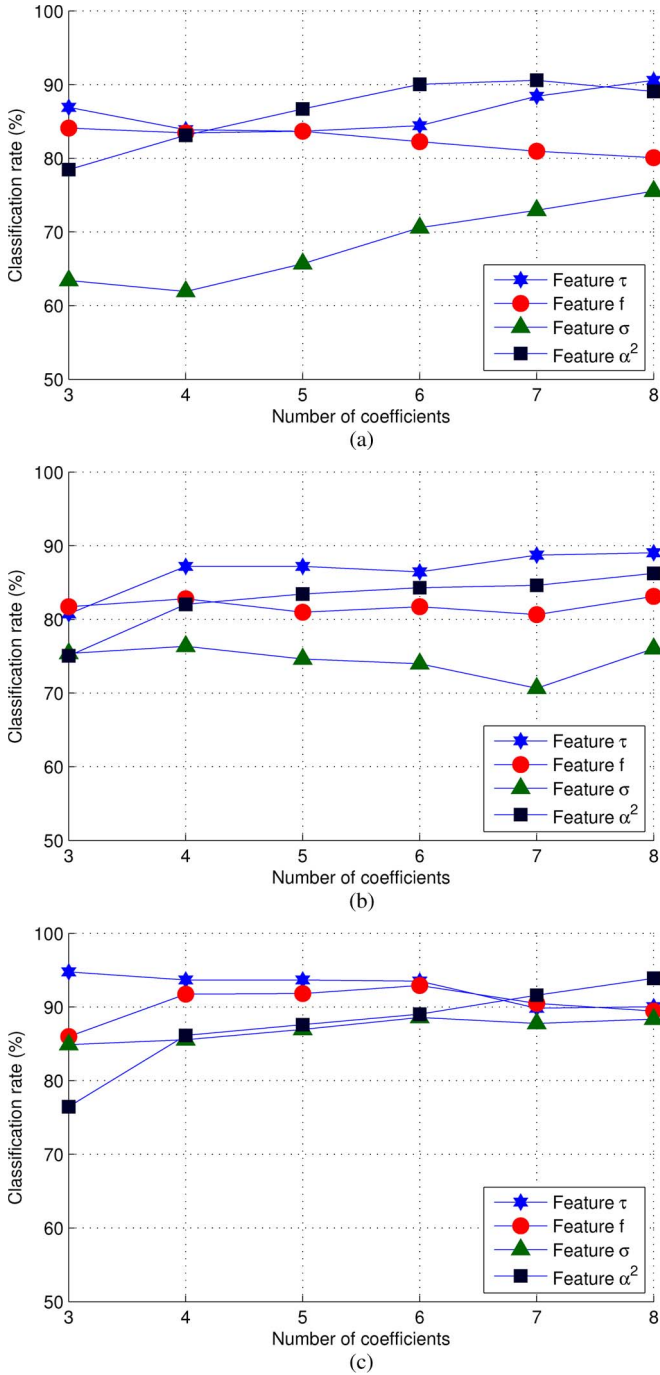


Fig. 6. Classification performance on the three data sets using one single feature.

TABLE V
CLASSIFICATION RATES (IN PERCENT) OF THE COMPOSITE FEATURE VECTOR ON THE THREE DATA SETS

Number of coefficients	Set 1	Set 2	Set 3
3	94.2±1.5	91.7±1.8	99.3±0.5
5	93.8±1.6	91.2±1.8	97.6±0.8
7	94.3±1.5	93.0±1.6	96.6±1.0

and Set 3, respectively. The composite feature vector is also evaluated on a combined data set comprising of Set 1, Set 2, and Set 3. The classification rate is 94.5% with three coefficients from each feature set.

TABLE VI
CLASSIFICATION RATES (IN PERCENT) OF SVMs WITH FEATURES EXTRACTED USING THE PROPOSED METHOD, STFT, AND WAVELETS

Features	Set 1	Set 2	Set 3
Proposed composite feature vector	94.2±1.5	91.7±1.8	99.3±0.5
STFT	87.1±2.1	92.2±1.7	99.2±0.5
Wavelet	78.1±2.7	72.3±2.9	95.6±1.1

C. Comparison With Other Features

In this section, we compare the classification performance of the sparse decomposition features with wavelet and STFT features. The sparse decomposition feature vector consists of the first three elements from each feature vector τ , f , σ , and α^2 . For the wavelet features, the discrete-time wavelet transform is applied to each preprocessed GPR trace. The wavelet coefficients are then normalized by the mean 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 pairwise SVM configuration is used with all feature vectors, and fivefold cross-validation is used to compute the overall classification rates. Table VI presents a comparison of 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 for signal classification.

V. CONCLUSION

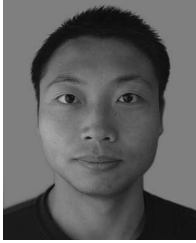
In this paper, we have proposed an adaptive sparse decomposition for GPR signal analysis and classification. It employs an overcomplete Gabor dictionary that is dynamically refined during the sparse decomposition. Furthermore, the proposed adaptive signal decomposition was found to be very effective for both signal representation and classification. Compared to the discrete wavelet transform and K-SVD, the proposed SR 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

The Windmill Islands data were obtained from the Australian Antarctic Data Center (International Directory Network Node Antarctic Master Directory/Australia), Australian Antarctic Division, Commonwealth of Australia. The Wollongong railway ground penetrating radar data were collected as part of the Rail CRC-AT5 project, sponsored by Cooperative Research Center Rail for Innovation.

REFERENCES

- [1] M. Skolnik, Ed., *Radar Handbook*, 3rd ed. New York: McGraw-Hill, 2008.
- [2] J. M. Reynolds, *An Introduction to Applied and Environmental Geophysics*. New York: Wiley, 1996.
- [3] H. M. Jol, Ed., *Ground Penetrating Radar Theory and Applications*, 1st ed. Amsterdam, The Netherlands: Elsevier Science, 2009.
- [4] J. Francke, "Applications of GPR in mineral resource evaluations," in *Proc. 13th Int. Conf. Ground Penetrating Radar*, 2010, pp. 1–5.
- [5] A. Denis, F. Huneau, S. Hoerl, and A. Salomon, "GPR data processing for fractures and flakes detection in sandstone," *J. Appl. Geophys.*, vol. 68, no. 2, pp. 282–288, Jun. 2009.
- [6] A. Yarovoy and M. J. Harry, "Landmine and unexploded ordnance detection and classification with ground penetrating radar," in *Ground Penetrating Radar Theory and Applications*. Amsterdam, The Netherlands: Elsevier, 2009, pp. 445–478.
- [7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [8] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [9] O. G. Guleryuz, "Weighted averaging for denoising with overcomplete dictionaries," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 3020–3034, Dec. 2007.
- [10] M. Rostami, O. Michailovich, and Z. Wang, "Image deblurring using derivative compressed sensing for optical imaging application," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3139–3149, Jul. 2012.
- [11] L. Peotta, L. Granai, and P. Vanderheyne, "Image compression using an edge adapted redundant dictionary and wavelets," *Signal Process.*, vol. 86, no. 3, pp. 444–456, Mar. 2006.
- [12] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
- [13] A. Castrodad, Z. Xing, J. B. Greer, E. Bosch, L. Carin, and G. Sapiro, "Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4263–4281, Nov. 2011.
- [14] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [15] Z. Tang, G. Blacquière, and G. Leus, "Aliasing-free wideband beamforming using sparse signal representation," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3464–3469, Jul. 2011.
- [16] J. Ma and F. X. Le Dimet, "Deblurring from highly incomplete measurements for remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 792–802, Mar. 2009.
- [17] J. L. Starck, F. Murtagh, and J. M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
- [18] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York: Springer-Verlag, 2010.
- [19] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. IEEE Radar Conf.*, 2007, pp. 128–133.
- [20] A. C. Gurbuz, J. H. McClellan, and W. R. Scott, "A compressive sensing data acquisition and imaging method for stepped frequency GPRs," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2640–2650, Jul. 2009.
- [21] A. C. Gurbuz, J. H. McClellan, and W. R. Scott, Jr., "Compressive sensing for subsurface imaging using ground penetrating radar," *Signal Process.*, vol. 89, no. 10, pp. 1959–1972, Oct. 2009.
- [22] M. G. Amin, Ed., *Through-the-Wall Radar Imaging*, 1st ed. Boca Raton, FL: CRC Press, 2010.
- [23] I. Kyriakides, "Adaptive compressive sensing and processing of delay-Doppler radar waveforms," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 730–739, Feb. 2012.
- [24] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.
- [25] A. C. Gurbuz, J. H. McClellan, and W. R. Scott, "Compressive sensing for GPR imaging," in *Proc. 41st Asilomar Conf. Signals, Syst. Comput.*, 2007, pp. 2223–2227.
- [26] A. C. Gurbuz, J. H. McClellan, and W. R. Scott, "GPR imaging using compressed measurements," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, vol. 2, pp. II-13–II-16.
- [27] L. Qu and T. Yang, "Investigation of air/ground reflection and antenna beamwidth for compressive sensing SFCW GPR migration imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3143–3149, Aug. 2012.
- [28] A. B. Suksmono, E. Bharata, A. A. Lestari, A. G. Yarovoy, and L. P. Ligthart, "Compressive stepped-frequency continuous-wave ground-penetrating radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 665–669, Oct. 2010.
- [29] F. Soldovieri, R. Solimene, L. Lo Monte, M. Bavusi, and A. Loperte, "Sparse reconstruction from GPR data with applications to rebar detection," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 3, pp. 1070–1079, Mar. 2011.
- [30] Y.-S. Yoon and M. G. Amin, "Imaging of behind the wall targets using wideband beamforming with compressive sensing," in *Proc. IEEE/SP 15th Workshop Statist. Signal Process.*, 2009, pp. 93–96.
- [31] J. Yang, A. Bouzerdoum, and M. G. Amin, "Multi-view through-the-wall radar imaging using compressed sensing," in *Proc. 18th Eur. Signal Process. Conf.*, Aalborg, Denmark, 2010, pp. 1429–1433.
- [32] Q. Huang, L. Qu, B. Wu, and G. Fang, "UWB throughwall imaging based on compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1408–1415, Mar. 2010.
- [33] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.
- [34] M. R. McClure and L. Carin, "Matching pursuits with a wave-based dictionary," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 2912–2927, Dec. 1997.
- [35] U. Spagnolini and V. Rampa, "Multitarget detection/tracking for monostatic ground penetrating radar: Application to pavement profiling," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 383–394, Jan. 1999.
- [36] J. Irving and R. Knight, "Numerical modeling of ground-penetrating radar in 2-D using MATLAB," *Comput. Geosci.*, vol. 32, no. 9, pp. 1247–1258, Nov. 2006.
- [37] W. Shao, A. Bouzerdoum, and S. L. Phung, "Sparse signal decomposition for ground penetrating radar," in *Proc. IEEE Radar Conf.*, Kansas City, MO, 2011, pp. 453–457.
- [38] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [39] B. J. Allred, J. J. Daniels, and M. R. Ehsani, Eds., *Handbook of Agricultural Geophysics*. Hoboken, NJ: CRC Press, 2008.
- [40] J. Pettersson, *Ground Penetrating Radar Data From Windmill Islands—Miscellaneous data*. Tasmania, Australia: Aust. Antarctic Data Centre, 2006. [Online]. Available: <http://data.aad.gov.au/aadc/metadata/>
- [41] W. Shao, A. Bouzerdoum, S. L. Phung, L. Su, B. Indraratna, and C. Rujikiatkamjorn, "Automatic classification of ground-penetrating-radar signals for railway-ballast assessment," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3961–3972, Oct. 2011.
- [42] O. Scherzer, Ed., *Handbook of Mathematical Methods in Imaging*, 1st ed. New York: Springer Science+BusinessMedia, 2011.
- [43] S. G. Mallat and Z. Zhifeng, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [44] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approx.*, vol. 13, no. 1, pp. 57–98, 1997.
- [45] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [46] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
- [47] E. Candès and J. Romberg, *ℓ_1 -magic: Recovery of Sparse Signals via Convex Programming*. Pasadena, CA: Caltech, 2005. [Online]. Available: <http://users.ece.gatech.edu/justin/l1magic/>
- [48] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming*. Natick, MA: CVX Research, Inc., 2011. [Online]. Available: <http://cvxr.com/cvx/>
- [49] O. Lopera, N. Milisavljevic, D. Daniels, and B. Macq, "Time-frequency domain signature analysis of GPR data for landmine identification," in *Proc. 4th Int. Workshop Adv. Ground Penetrating Radar*, 2007, pp. 159–162.
- [50] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
- [51] S. Abe, *Support Vector Machines for Pattern Classification*. New York: Springer-Verlag, 2005.
- [52] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2001.
- [53] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, JULY 1998.
- [54] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Boston, MA: Morgan Kaufman, 2005.



Wenbin Shao received the B.Eng. degree in communication engineering from Northwestern Polytechnical University, Xi'an, China, in 2003 and the M.Eng. degree from the University of Wollongong, Wollongong, Australia, in 2007, where he is currently working toward the Ph.D. degree in electrical engineering.



Son Lam Phung (M'02) received the B.Eng. degree with first-class honors and the Ph.D. degree in computer engineering from Edith Cowan University, Perth, Australia, in 1999 and 2003, respectively. He received the University and Faculty Medals in 2000.

He is currently a Senior Lecturer with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, Australia. His general research interests are in the areas of image and signal processing, neural networks, pattern recognition, and machine

learning.



Abdesselam Bouzerdoum (M'89–SM'03) received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Washington, Seattle.

In 1991, he joined The University of Adelaide, Adelaide, Australia, and in 1998, he was appointed an Associate Professor with Edith Cowan University, Perth, Australia. Since 2004, he has been with the University of Wollongong, Wollongong, Australia, as Professor of Computer Engineering, where he also served as Head of School of Electrical, Computer and Telecommunications Engineering from 2004 to

2006 and Associate Dean (Research), with the Faculty of Informatics, from 2007 to 2013. He held a number of visiting appointments at several universities and research institutes, including Institut Galilée (Université Paris 13), Laboratoire d'Analyse et d'Architectures des Systèmes/Centre National de la Recherche Scientifique, Toulouse, France, The Hong Kong University of Science and Technology, Kowloon, Hong Kong, and Villanova University, Villanova, PA. From 2009 to 2011, he served as a Member of the Australian Research Council College of Experts, and was the Deputy Chair of the Engineering, Mathematics and Informatics panel from 2010 to 2011.

Dr. Bouzerdoum was the recipient of numerous awards and prizes; the most notable are the Eureka Prize for Outstanding Science in Support of Defence or National Security in 2011, the Chester Sall Award in 2005, and a Distinguished Researcher Award (Chercheur de Haut Niveau) from the French Ministry of Research in 2001. He has published over 280 technical articles and graduated many Ph.D. and Masters students. From 1999 to 2006, he served as Associate Editor of IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS. He is currently an Associate Editor of three international journals and a member of the governing board of the Asia Pacific Neural Network Assembly. He is a member of Optical Society of America and International Neural Network Society.