2019

GPR Target Detection by Joint Sparse and Low-Rank Matrix Decomposition

Fok Hing Chi Tivive
University of Wollongong, tivive@uow.edu.au

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

Canicious Abeynayake
Defence Science and Technology Group

GPR Target Detection by Joint Sparse and Low-Rank Matrix Decomposition

Abstract

Ground penetrating radar (GPR) uses electromagnetic waves to image, locate, and identify changes in electric and magnetic properties in the ground. The received signal comprises not only the target echoes but also strong reflections from the rough, uneven ground surface, which impair subsurface inspections and visualization of buried objects. In this paper, a background clutter mitigation and target detection method using low-rank and sparse priors is proposed for GPR data. The radar signal is decomposed into the sum of a low-rank component and a sparse component, plus noise. The low-rank component captures the ground surface reflections and background clutter, whereas the sparse component contains the target reflections. The effectiveness of the proposed method is evaluated on real radar signals collected from buried landmines and improvised explosive devices. The experimental results show that the proposed method successfully removes the background clutter and estimates the target signals.

Disciplines

Engineering | Science and Technology Studies

Publication Details

F. Tivive, A. Bouzerdoum & C. Abeynayake, "GPR Target Detection by Joint Sparse and Low-Rank Matrix Decomposition," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, (5) pp. 2583-2595, 2019.

GPR Target Detection by Joint Sparse and Low-Rank Matrix Decomposition

Fok Hing Chi Tivive[1], *Member, IEEE,* Abdesselam Bouzerdoum[1,2], *Senior Member, IEEE,* and Canicious Abeynayake[3]

***Abstract*—Ground penetrating radar uses electromagnetic waves to image, locate, and identify changes in electric and magnetic properties in the ground. The received signal comprises not only the target echoes, but also strong reflections from the rough, uneven ground surface, which impair subsurface inspections and visualization of buried objects. In this paper, a background clutter mitigation and target detection method using low-rank and sparse priors is proposed for ground penetrating radar data. The radar signal is decomposed into the sum of a low-rank component and a sparse component, plus noise. The low-rank component captures the ground surface reflections and background clutter, whereas the sparse component contains the target reflections. The effectiveness of the proposed method is evaluated on real radar signals collected from buried landmines and improvised explosive devices. The experimental results show that the proposed method successfully removes the background clutter and estimates the target signals.**

***Index Terms*—Ground penetrating radar, clutter mitigation, low-rank and sparse priors, low-rank representation, target detection.**

I. Introduction

Ground penetrating radar (GPR) is a non-intrusive and non-destructive sensing modality which has been used in many civilian and military applications, including mine detection [1], bridge and tunnel assessment [2], utility mapping [3], ballast assessment [4], and void detection [5]. For the detection of landmines and improvised explosive devices, the GPR system transmits electromagnetic waves to sense electrical inhomogeneities exhibited by the buried targets. In addition to the target reflections, radar returns from the ground surface and subsurface soil layering are also received by the antenna array. The ground surface reflections, which are stronger than the target reflections, render target detection very difficult, or even impossible. Therefore, a major research effort in GPR has been devoted to developing novel approaches for suppressing the background clutter, including the ground surface reflections, background scattering, and background noise [6]–[30].

F. H. C. Tivive is with the [1]School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Northfields Avenue, Wollongong NSW 2522, Australia (Email: tivive@uow.edu.au).

A. Bouzerdoum is with the [1]School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Northfields Avenue, Wollongong NSW 2522, Australia; [2]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar (Email: a.bouzerdoum@uow.edu.au).

C. Abeynayake is with the [3]Weapons and Combat Systems Division, Defence Science and Technology Group, Edinburgh, SA 5111, Australia (Email: Canicious.Abeynayake@dsto.defence.gov.au).

The approaches for removing background clutter in GPR data can be grouped into five different categories: time gating, filtering, parametric modeling, and subspace decomposition. Time gating is a technique to remove the portion of the signal before a certain time or range that comprises the strong clutter since the direct return from the transmitter and the signal backscattered from the ground surface arrive much earlier than the target return [8]. When the target is buried near the ground surface, the target response overlaps with the strong ground reflections; therefore, it is very difficult to determine an appropriate time window that covers only the background clutter. Solimene *et al.* employed statistical entropy to design the time window for removing the ground surface reflections and background noise [9]. They firstly assumed that the background signal has similar characteristics across the antenna array and the target signals consist of different delayed pulses. Then, the window is determined by identifying the time bin whose entropy is greater than a threshold, where the threshold is determined as a scale factor of the length of the radar trace. However, the entropy-based time gating method is effective only when the antenna array is placed parallel to the ground surface.

Several filtering methods have been developed for GPR clutter removal. One simple filtering technique is mean subtraction, where an estimate of the average signal trace is obtained by calculating the mean of a number of received signals. Median filtering was also considered in [10] to cope with noisy signal traces and outliers. These simple filtering techniques work well when there is not much overlapping between the target and the ground surface returns, and the ground surface is smooth. More sophisticated filtering methods were proposed to cope with these cases where overlap occurs. In [11], a digital high-pass filter was developed based on the observations that the clutter and the buried objects appear, respectively, as horizontal bands and hyperbolas in the GPR data matrix, known as the B-scan. The cutoff frequency of the digital filter was determined from the spectrum of the clutter segment. In [12], the symmetry filtering technique was proposed based on the assumption that the target response has a symmetric shape and the background reflections appear to be random and unsymmetric. Other filtering methods first transform the GPR data into another domain (e.g., wavelet or curvelet) and then remove the coefficients carrying the clutter. In [13], the skewness statistic was used to determine the set of wavelet coefficients containing the background clutter, which was assumed to be white Gaussian. In [14], the B-scan was converted into a set of curvelet coefficients. The coefficients

in the coarsest layer and those corresponding to a zero slope curvelet were set to zeros. However, replacing the curvelet coefficients in the coarsest layer by zeros also removes some of the target reflections, particularly those having horizontal shape.

Instead of filtering out or gating the clutter signal, it can be modeled from the received signals. To this end, a number of model-based approaches have been proposed [15]–[17]. Brunzell introduced a least squares method to estimate the background signal [15]. Merwe and Gupta modeled the clutter signal as a superposition of damped complex exponentials and proposed an iterative technique to estimate their parameters [16]. Chan *et al.* developed a two-sided linear prediction technique to determine the background signal from the B-scan [17]. These model-based methods, however, required a portion of the B-scan containing clutter only to estimate the model parameters. On the other hand, subspace approaches model the ground surface reflections as a low-rank subspace based on the observations that the ground surface reflections are stronger than the target reflections, and they are highly correlated among the signals received across the antenna array. Therefore, several subspace decomposition techniques, such as singular value decomposition (SVD) [18]–[21], principal component analysis (PCA) [22]–[24], and independent component analysis (ICA) [25]–[28], have been employed to decompose the radar signal into three different components: clutter, target, and noise. The SVD and PCA based techniques assume the clutter lies in a subspace spanned by the dominant eigencomponents, whereas the ICA based approach considers the clutter to be captured by independent components having Gaussian characteristics. The drawback of these subspace decomposition methods is how to differentiate between the components spanning the clutter, target, and noise subspaces. Riaz *et al.* assumed the first dominant component to span the clutter subspace and applied minimum description length (MDL) or Akaike information criterion (AIC) to separate the target and noise subspaces [20]. However, the rank of the background clutter subspace can be greater than one in practical applications, due to the inhomogeneity of the soil and roughness of the ground surface. Most recently, robust principal component analysis (RPCA), which is an extension of the standard PCA to cope with grossly corrupted data, was employed for GPR anomaly detection [29]. In this technique, the radar signal undergoes a series of preprocessing steps such as signal alignment, haircutting, and data transformation before signal decomposition. In [30], different optimization algorithms for solving the RPCA problem were evaluated for removing clutter in GPR data. Among the tested optimization algorithms, the principal component pursuit by alternating directions and the ℓ_1-norm filtering techniques were shown to achieve superior results.

This paper introduces a technique to estimate the background clutter and the target signal from GPR traces, using low-rank and sparse priors to decompose the received GPR signals into the sum of a low-rank matrix and a sparse matrix. A joint low-rank and sparse (JLRS) model is formulated to extract a low-rank representation of the background clutter and a sparse representation of the target signal. The proposed model is solved using an optimization algorithm based on the alternating direction method of multipliers (ADMM). Unlike RPCA, which is concerned solely with recovering a low-rank structure of the input data, the proposed method employs additional sparsity constraints to estimate a sparse matrix containing the target reflections, in addition to a low-rank matrix for background clutter. Compared to the low-rank based method described in [29] that preprocessed the GPR data before signal decomposition, the proposed JLRS method employs analysis or synthesis priors to determine the low-rank and sparse representations. Moreover, the proposed JLRS model takes into account the noise of the GPR signal in the estimation the low-rank and sparse components.

The remainder of this paper is organized as follows. Section II presents the formulation of the background clutter removal and target detection as a joint low-rank and sparse optimization problem, followed by existing low-rank and sparse matrix decomposition methods. Section III describes the proposed JLRS models, their optimization algorithms, and the target detection scheme. Section IV presents the experimental results and discussion, and Section V gives the conclusion.

II. Problem Formulation

GPR signals collected by the radar system along either the cross-track or down-track direction are stacked as columns to form the B-scan, denoted by $Y = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N] \in \mathbb{R}^{M \times N}$, where M is the number of depth or time bins and N is the number of antenna locations. The proposed JLRS method is based on the assumption that the received ith radar signal $\boldsymbol{y}_i \in \mathbb{R}^{M \times 1}$ $(i = 1, ..., N)$ is composed of the background clutter $\boldsymbol{l}_i$, the target return $\boldsymbol{s}_i$, and noise $\boldsymbol{e}_i$. The background clutter of the B-scan, which has similar characteristics between the radar traces, forms a low-rank matrix $L = [\boldsymbol{l}_1, \ldots, \boldsymbol{l}_N]$, whereas the target signal forms a sparse matrix $S = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N]$. Therefore, the B-scan can be expressed as

$$Y = L + S + E, \tag{1}$$

where $E = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N]$ is an error matrix containing noise. Robust principal component analysis (RPCA) and low-rank representation (LRR) are the two common approaches for decomposing a data matrix into low-rank and sparse matrices. Their mathematical models are described in Subsections II-A and II-B.

A. Robust Principal Component Analysis

RPCA was proposed to circumvent the drawback of classical PCA, which is sensitive to outliers. Its aim is to find a low-rank structure in high dimensional data by solving the following optimization problem:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad \text{s.t. } Y = L + S, \tag{2}$$

where $||\cdot||_*$ denotes the nuclear norm of the matrix argument (i.e., the sum of its singular values), $||\cdot||_1$ denotes the ℓ_1-norm (i.e., the sum of the absolute values of matrix entries), and λ is a positive regularization parameter. Under certain noise sparsity and rank upper-bound assumptions, the matrix L can be exactly recovered from Y as long as S is sufficiently sparse [31].

B. Low Rank Representation

Contrary to RPCA, which performs matrix recovery under the assumption that the underlying data structure is a single low-rank subspace, the LRR method focuses on finding the lowest-rank representation of the data by solving the following optimization problem [32]:

$$\min_{Z,S} \|Z\|_* + \lambda \|S\|_{2,1} \quad \text{s.t. } Y = \Theta Z + S, \tag{3}$$

where $\|S\|_{2,1} = \sum_{i=1}^{N} \|\boldsymbol{s}_i\|_2$ is the mixed $\ell_{2,1}$-norm used to encourage the matrix $S = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N]$ to be column-sparse, Θ is a dictionary that linearly spans the data space, and Z denotes the lowest-rank representation of the data Y with respect to the dictionary Θ.

The RPCA and LRR models given in (2) and (3) mainly focus on estimating the low-rank matrix by minimizing the nuclear norm and performing error correction by minimizing the ℓ_1-norm or $\ell_{2,1}$-norm. The next section presents the proposed JLRS model, which estimates not only a low-rank matrix, but also a sparse matrix while taking into account the signal noise.

III. Proposed Joint Low Rank and Sparse Method

This section presents two alternative joint low-rank and sparse signal representations, which are used to model the background clutter and the target signal in the B-scan. Then, optimization techniques based on ADMM and Bayesian theory are developed to solve the JLRS models and tune their regularization parameters. Finally, a constant false alarm rate (CFAR) detector is employed to detect and localize the target signals in the estimated sparse matrix.

In the following, two signal decomposition techniques are proposed for estimating the low-rank and sparse matrices: joint low-rank and sparse representation using synthesis prior (JLRS-SP) and joint low-rank and sparse representation using analysis prior (JLRS-AP). In LRR, the synthesis model is applied to determine the low-rank representation Z from which the low-rank matrix is computed as $L = \Theta Z$. In the proposed JLRS-SP, on the other hand, the synthesis model is applied to both low-rank and sparse representations, while taking into account the noise of the B-scan. The low-rank matrix is obtained similarly to LRR, whereas the sparse matrix is computed as $S = \Omega X$, where Ω is a sparse synthesis dictionary and X is the estimated sparse representation. Thus, the JLRS-SP model can be formulated as

$$\min_{Z,X} \|Z\|_* + \lambda_1 \|X\|_1 + \lambda_2 \|X\|_{2,1} \text{ s.t. } Y = \Theta Z + \Omega X + E, \tag{4}$$

where λ_1 and λ_2 are regularization parameters and E is the noise in the B-scan signal. The JLRS-SP model employs both ℓ_1-norm and mixed $\ell_{2,1}$-norm to promote sparsity and remove any columns that do not contain target information in the matrix S.

The second proposed method, JLRS-AP, is based on the analysis model, which is the counterpart of the synthesis model. Let Ψ denote the sparse analysis operator, which is the sparse analysis dictionary defined as $\Psi = \Omega^{\dagger}$, where † is the pseudo-inverse operator, and let Φ be the low-rank analysis operator, $\Phi = \Theta^{\dagger}$. In the analysis model, the matrix ΨS is expected to be sparse and the matrix $Z = \Phi L$ to be low-rank. Using the analysis prior, the JLRS-AP model can be written as

$$\min_{L,S} \|\Phi L\|_* + \lambda_1 \|\Psi S\|_1 + \lambda_2 \|\Psi S\|_{2,1} \text{ s.t. } Y = L + S + E. \tag{5}$$

Note that the JLRS-SP and JLRS-AP models are equivalent when the sparse synthesis dictionary is square and invertible, i.e., $\Omega = \Psi^{-1}$ [33]. However, when the sparse synthesis dictionary is overcomplete or redundant, the two models produce different results.

In both models, the M rows of the noise matrix E, $\boldsymbol{e}_m$ $(m = 1, \ldots, M)$, are assumed to be i.i.d. random vectors, following a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \Sigma)$, where Σ is the covariance matrix. Let vec$(\cdot)$ denote the vectorization operator stacking the columns (or rows) of a matrix into a column (or row) vector. The distribution of the error vector, vec(E), is given by

$$\text{vec}(E) = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_M]^T \sim \mathcal{N}_{MN}(\mathbf{1}_M \otimes \mathbf{0}, \Lambda), \tag{6}$$

where ^{T} is the transpose operator, $\otimes$ is the Kronecker product, $\Lambda = I_M \otimes \Sigma$, $\mathbf{1}_M$ is the M-dimensional column vector of ones, and I_M is the $M \times M$ identity matrix. The noise matrix E can be characterized by the matrix-variate one, i.e.,

$$E \sim \mathcal{N}_{M,N}(\mathbf{1}_M \mathbf{0}^T, I_M, \Sigma). \tag{7}$$

The likelihood function $\mathcal{L}(L, S, \Sigma; Y)$ is then given by

$$\mathcal{L}(L, S, \Sigma; Y) = (2\pi)^{-\frac{MN}{2}} |\Sigma|^{-\frac{M}{2}} |I_M|^{-\frac{N}{2}} \exp\Big(-\frac{1}{2} \text{tr}(I_M (Y - L - S) \Sigma^{-1} (Y - L - S)^T) \Big), \tag{8}$$

where $|\cdot|$ is the determinant of a square matrix and tr$(\cdot)$ is the trace of a matrix. Given an estimated of the noise covariance matrix Σ, the matrices L and S can be determined by maximizing the log-likelihood function:

$$\ln \mathcal{L}(L, S, \Sigma; Y) = -\frac{MN}{2} \ln(2\pi) - \frac{N}{2} \ln(I_M) - \frac{M}{2} \ln(\Sigma) - \frac{1}{2} \text{SSE}(L, S), \tag{9}$$

where $\text{SSE}(L, S) = \text{tr}[I_M (Y - L - S) \Sigma^{-1} (Y - L - S)^T]$ is the sum of squared errors. Maximizing the log-likelihood is equivalent to minimizing the SSE. Therefore, in our case, the optimization problem to be solved can be stated as follows:

$$\min_{L,S} \quad \frac{1}{2} \text{tr}[(Y - L - S) \Sigma^{-1} (Y - L - S)^T]. \tag{10}$$

By combining the low-rank and sparse regularization terms given in (4) with (10), the JLRS-SP model can be reformulated as

$$\min_{Z,X} \|Z\|_* + \lambda_1 \|X\|_1 + \lambda_2 \|X\|_{2,1} + \frac{\lambda_3}{2} \text{tr}[(Y - \Theta Z - \Omega X) \Sigma^{-1} (Y - \Theta Z - \Omega X)^T]. \tag{11}$$

Similarly, the JLRS-AP model can be rewritten as

$$\min_{L,S} \|\Phi L\|_* + \lambda_1 \|\Psi S\|_1 + \lambda_2 \|\Psi S\|_{2,1} + \frac{\lambda_3}{2}\text{tr}[(Y-L-S)\Sigma^{-1}(Y-L-S)^T]. \quad (12)$$

Both (11) and (12) are constrained optimization problems with non-smooth regularization terms. Therefore, the augmented Lagrangian multiplier (ALM) described in [34], which is a variant of ADMM, is applied to solve these two problems.

A. Optimization Techniques for the Proposed JLRS Models

Several algorithms have been proposed to solve low-rank optimization problems, such as singular value thresholding [35], accelerated proximal gradient [36], Split-Bregman method [37], and augmented Lagrangian multiplier (ALM) [34]. In particular, ALM has received considerable attention due to its simple form and decoupling of variables. It has been used in compressed sensing [38], [39], image restoration and reconstruction [40], [41], and matrix completion and recovery [34], [42]. ALM is often used to solve convex, non-smooth objective functions with linear constraints. It updates the variables alternately by minimizing the augmented Lagrangian function. Lin *et al.* [34] proposed two ALM algorithms: exact ALM and inexact ALM. These two algorithms are faster than the singular value thresholding and accelerated proximal gradient techniques and have Q-linear convergence speed [34]. The difference between them is that inexact ALM is less computation intensive than the exact ALM since it does not need to solve each sub-problem exactly so long as each update of the variables converges to the optimal solution of the problem; proofs of their convergence are given in [34] and [43]. In this paper, two optimization methods based on the inexact ALM algorithm are proposed for solving Problems (11) and (12).

1) JLRS-SP Optimization Technique: By introducing two auxiliary variables F and G to make (11) separable, the JLRS-SP model can be formulated as a constrained optimization problem:

$$\min_{F,Z,G,X} \|F\|_* + \lambda_1 \|G\|_1 + \lambda_2 \|G\|_{2,1} + \frac{\lambda_3}{2}\text{tr}[(Y-\Theta Z-\Omega X)\Sigma^{-1}(Y-\Theta Z-\Omega X)^T] \quad \text{s.t.} \quad F=Z,\ G=X. \quad (13)$$

Combining the above objective function with the constraints yields the following augmented Lagrangian function:

$$\mathcal{L}(F,Z,G,X,B,C) = \|F\|_* + \lambda_1 \|G\|_1 + \lambda_2 \|G\|_{2,1} + \frac{\lambda_3}{2}\text{tr}[(Y-\Theta Z-\Omega X)\Sigma^{-1}(Y-\Theta Z-\Omega X)^T] + \langle B, Z-F\rangle + \langle C, X-G\rangle + \frac{\mu}{2}\big[\big\|Z-F\big\|_F^2 + \big\|X-G\big\|_F^2\big], \quad (14)$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product between two matrices, B and C are Lagrange multipliers, and $\mu > 0$ is a penalty parameter. By simplifying the last four terms (see Appendix for more details), the augmented Lagrangian function (14) can be concisely rewritten as

$$\mathcal{L}(F,Z,G,X,B,C) = \|F\|_* + \lambda_1 \|G\|_1 + \lambda_2 \|G\|_{2,1} + \frac{\lambda_3}{2}\text{tr}[(Y-\Theta Z-\Omega X)\Sigma^{-1}(Y-\Theta Z-\Omega X)^T] + \frac{\mu}{2}\big\|Z-F+\frac{B}{\mu}\big\|_F^2 + \frac{\mu}{2}\big\|X-G+\frac{C}{\mu}\big\|_F^2 - \frac{1}{2\mu}\big[\big\|B\big\|_F^2 + \big\|C\big\|_F^2\big]. \quad (15)$$

To solve for the variables F, Z, G, and X, the ALM algorithm updates each variable alternately by minimizing the augmented Lagrangian function, while keeping the other variables fixed. Therefore, Problem (15) can be decomposed into the following subproblems:

$$F^{k+1} = \arg\min_F \|F\|_* + \frac{\mu^k}{2}\big\|Z^k - F + \frac{B^k}{\mu^k}\big\|_F^2, \quad (16)$$

$$Z^{k+1} = \arg\min_Z \frac{\lambda_3}{2}\text{tr}[(Y-\Theta Z-\Omega X^k)\Sigma^{-1}(Y-\Theta Z-\Omega X^k)^T] + \frac{\mu^k}{2}\big\|Z - F^{k+1} + \frac{B^k}{\mu^k}\big\|_F^2, \quad (17)$$

$$X^{k+1} = \arg\min_X \frac{\lambda_3}{2}\text{tr}[(Y-\Theta Z^{k+1}-\Omega X)\Sigma^{-1}(Y-\Theta Z^{k+1}-\Omega X)^T] + \frac{\mu^k}{2}\big\|X - G^k + \frac{C^k}{\mu^k}\big\|_F^2, \quad (18)$$

$$G^{k+1} = \arg\min_G \lambda_1 \|G\|_1 + \lambda_2 \|G\|_{2,1} + \frac{\mu^k}{2}\big\|X^{k+1} - G + \frac{C^k}{\mu^k}\big\|_F^2. \quad (19)$$

$$B^{k+1} = B^k + \mu^k(Z^{k+1} - F^{k+1}), \quad (20)$$

$$C^{k+1} = C^k + \mu^k(X^{k+1} - G^{k+1}). \quad (21)$$

Subproblem (16) is a least squares problem regularized by a nuclear norm penalty. It can be efficiently solved using singular value shrinkage [44], [45]. Let $\mathcal{T}(a,\beta)$ denote the shrinkage operator

$$\mathcal{T}(a,\beta) = \text{sgn}(a)\max(|a|-\beta, 0). \quad (22)$$

Such shrinkage operator is applied entrywise to vectors as well as matrices. The minimization of Subproblem (16) can be performed by applying the shrinkage operator to the singular values as follows:

$$\text{SVD}\Big(Z^k + B^k/\mu^k\Big) = UDV^T, \quad (23)$$

$$F^{k+1} = U\mathcal{T}\big(D, 1/\mu^k\big)V^T. \quad (24)$$

where U and V are unitary matrices, and D is a diagonal matrix of singular values. Subproblem (17) is a least squares problem, which can be solved using a conjugate gradient method. Differentiating the right hand side of (17) with respect to Z and setting the result equal to zero, the following Sylvester equation is obtained:

$$\lambda_3\Theta^T\Theta Z + Z(\mu^k\Sigma) = \lambda_3\Theta^T(Y-\Omega X^k) + \mu^k(F^{k+1} - \frac{B^k}{\mu^k})\Sigma \quad (25)$$

Let $\mathrm{mat}(\cdot)$ denote the operator reshaping a column vector of MN elements into a $M \times N$ matrix. The solution of Subproblem (25) is given by

$$Z^{k+1} = \mathrm{mat}\Big([I \otimes (\lambda_3 \Theta^T \Theta) + \mu^k \Sigma^T \otimes I]^{-1} \mathrm{vec}(\lambda_3 \Theta^T (Y - \Omega X^k) + \mu^k (F^{k+1} - \frac{B^k}{\mu^k})\Sigma)\Big), \quad (26)$$

where I is the identity matrix. Similarly, Subproblem (18) has the following solution:

$$X^{k+1} = \mathrm{mat}\Big([I \otimes (\lambda_3 \Omega^T \Omega) + \mu^k \Sigma^T \otimes I]^{-1} \mathrm{vec}(\lambda_3 \Omega^T (Y - \Theta Z^{k+1}) + \mu^k (G^k - \frac{C^k}{\mu^k})\Sigma)\Big). \quad (27)$$

The last Subproblem (19) can be solved using a generalized shrinkage operator [37], defined column-wise by

$$\mathcal{G}(\boldsymbol{g}_i, \gamma_1, \gamma_2) = \frac{\mathcal{T}(\boldsymbol{g}_i, \gamma_1)}{1 + \big[\gamma_2 / \mathcal{T}\big(\|\mathcal{T}(\boldsymbol{g}_i, \gamma_1)\|_2, \gamma_2\big)\big]} \quad \forall i, \quad (28)$$

where $\boldsymbol{g}_i$ is the ith column of G, and $\gamma_1 = \lambda_1/\mu^k$ and $\gamma_2 = \lambda_2/\mu^k$ are threshold values. Let $\boldsymbol{c}_i$ denote the ith column of C. Using the shrinkage operator given in (28), the update of the ith column of matrix G^{k+1} is expressed as

$$\boldsymbol{g}_i^{k+1} = \mathcal{G}\big(\boldsymbol{x}_i^{k+1} + \boldsymbol{c}_i^k/\mu^k, \gamma_1, \gamma_2\big) \quad \text{for } i = 1, \ldots, N. \quad (29)$$

2) JLRS-AP Optimization Technique: To solve the JLRS-AP model given in (12) using the ALM algorithm, two auxiliary variables Q and R are introduced to the objective function, which can be expressed as

$$\begin{aligned} \min_{Q,L,S,R} \|Q\|_* + \lambda_1 \|R\|_1 + \lambda_2 \|R\|_{2,1} \\ + \frac{\lambda_3}{2}\mathrm{tr}[(Y - L - S)\Sigma^{-1}(Y - L - S)^T], \\ \text{s.t. } \Phi L = Q,\ \Psi S = R. \end{aligned} \quad (30)$$

The augmented Lagrangian function can be written as

$$\begin{aligned} \mathcal{L}(Q, L, S, R, \widehat{B}, \widehat{C}) = \|Q\|_* + \lambda_1 \|R\|_1 + \lambda_2 \|R\|_{2,1} \\ + \frac{\lambda_3}{2}\mathrm{tr}[(Y - L - S)\Sigma^{-1}(Y - L - S)^T] + \\ \frac{\mu}{2}\big(\big\|\Phi L - Q + \frac{\widehat{B}}{\mu}\big\|_F^2 + \big\|\Psi S - R + \frac{\widehat{C}}{\mu}\big\|_F^2\big), \end{aligned} \quad (31)$$

where $\widehat{B}$ and $\widehat{C}$ are Lagrange multipliers. Problem (31) can be divided into the following subproblems:

$$Q^{k+1} = \arg\min_Q \|Q\|_* + \frac{\mu^k}{2}\big\|\Phi L^k - Q + \frac{\widehat{B}^k}{\mu^k}\big\|_F^2, \quad (32)$$

$$L^{k+1} = \arg\min_L \frac{\lambda_3}{2}\mathrm{tr}[(Y - L - S^k)\Sigma^{-1}(Y - L - S^k)^T] + \frac{\mu^k}{2}\big\|\Phi L - Q^{k+1} + \frac{\widehat{B}^k}{\mu^k}\big\|_F^2, \quad (33)$$

$$S^{k+1} = \arg\min_S \frac{\lambda_3}{2}\mathrm{tr}[(Y - L^{k+1} - S)\Sigma^{-1}(Y - L^{k+1} - S)^T] + \frac{\mu^k}{2}\big\|\Psi S - R^k + \frac{\widehat{C}^k}{\mu^k}\big\|_F^2, \quad (34)$$

$$R^{k+1} = \arg\min_R \lambda_1 \|R\|_1 + \lambda_2 \|R\|_{2,1} + \frac{\mu^k}{2}\big\|\Psi S^{k+1} - R + \frac{\widehat{C}^k}{\mu^k}\big\|_F^2. \quad (35)$$

The solution to (32), using the singular value shrinkage operator, can be determined as

$$\begin{aligned} \mathrm{SVD}\Big(\Phi L^k + \widehat{B}^k/\mu^k\Big) &= UDV^T, & (36) \\ Q^{k+1} &= U\mathcal{T}\big(D, 1/\mu^k\big)V^T. & (37) \end{aligned}$$

The Sylvester equations obtained from solving Subproblems (33) and (34) are

$$(\mu^k \Phi^T \Phi)L + L(\lambda_3 \Sigma^{-1}) = \lambda_3 (Y - S^k)\Sigma^{-1} + \mu^k \Phi^T \Big(Q^{k+1} - \frac{\widehat{B}^k}{\mu^k}\Big) \quad (38)$$

and

$$(\mu^k \Psi^T \Psi)S + S(\lambda_3 \Sigma^{-1}) = \lambda_3 (Y - L^{k+1})\Sigma^{-1} + \mu^k \Psi^T \Big(R^k - \frac{\widehat{C}^k}{\mu^k}\Big). \quad (39)$$

Their solutions are given by

$$L^{k+1} = \mathrm{mat}\Big([I \otimes (\mu^k \Phi^T \Phi) + (\lambda_3 \Sigma^{-1})^T \otimes I]^{-1} \mathrm{vec}(\lambda_3 (Y - S^k)\Sigma^{-1} + \mu^k \Phi^T (Q^{k+1} - \frac{\widehat{B}^k}{\mu^k}))\Big) \quad (40)$$

and

$$S^{k+1} = \mathrm{mat}\Big([I \otimes (\mu^k \Psi^T \Psi) + (\lambda_3 \Sigma^{-1})^T \otimes I]^{-1} \mathrm{vec}(\lambda_3 (Y - L^{k+1})\Sigma^{-1} + \mu^k \Psi^T (R^k - \frac{\widehat{C}^k}{\mu^k}))\Big). \quad (41)$$

Subproblem (35) can be solved using the generalized shrinkage operator given in (28), where the ith column of R^{k+1} can be expressed as

$$\boldsymbol{r}_i^{k+1} = \mathcal{G}\big(\Psi \boldsymbol{s}_i^{k+1} + \widehat{\boldsymbol{c}}_i^k/\mu^k, \gamma_1, \gamma_2\big) \quad \text{for} \quad i = 1, \ldots, N. \quad (42)$$

Finally, the Lagrange multipliers $\widehat{B}$ and $\widehat{C}$ are updated as follows:

$$\widehat{B}^{k+1} = \widehat{B}^k + \mu^k(\Phi L^{k+1} - Q^{k+1}) \quad (43)$$

and

$$\widehat{C}^{k+1} = \widehat{C}^k + \mu^k(\Psi S^{k+1} - R^{k+1}). \quad (44)$$

The steps of the optimization methods for solving the JLRS-SP and JLRS-AP models are summarized, respectively, as Algorithm-1 and Algorithm-2 in the Appendix. In both algorithms, the variables L^0, S^0, Z^0, X^0 are initialized as follows. SVD is firstly applied to the transformed data ΦY or $\Theta^\dagger Y$, where $\Theta^\dagger$ is the pseudo-inverse of Θ. Then, the K

dominant singular vectors is determined by applying Otsu's threshold technique to the singular values, similar to [46]. Let $\mathbf{u}_i$ and $\mathbf{v}_i$ be the left and right ith singular vectors, respectively, and β_i be the ith singular value of ΦY. The matrices L^0 and S^0 are computed as $L^0 = \Phi^\dagger(\sum_{i=1}^{K} \beta_i \mathbf{u}_i \mathbf{v}_i^T)$ and $S^0 = Y - L_0$, respectively. Similarly, after SVD of $\Theta^\dagger Y$, the matrices Z^0 and X^0 are initialized as: $Z^0 = \sum_{i=1}^{K} \hat{\beta}_i \hat{\mathbf{u}}_i \hat{\mathbf{v}}_i^T$ and $X^0 = \Omega^\dagger(Y - \Theta Z^0)$, where $\hat{\mathbf{u}}_i$, $\hat{\mathbf{v}}_i$ and $\hat{\beta}_i$ are the ith singular vectors and singular value of $\Theta^\dagger Y$, respectively.

3) Computational Complexity: The proposed JLRS optimization methods solve several subproblems in order to estimate a low-rank matrix and a sparse matrix from the B-scan. The minimization of the nuclear norm and the solution of the Sylvester equation are the two most time-consuming steps. Suppose the number of atoms J in the dictionary is greater than the number of traces N in the B-scan of size $M \times N$, i.e., $M < N < J$. The minimization of the nuclear norm involves an SVD step that has a computational complexity of $\mathcal{O}(JN^2)$. The solutions of the Sylvester equations given by Eq. (26) and (27) for JLRS-SP require a matrix inversion, which has a computational complexity of $\mathcal{O}((JN)^3)$. For JLRS-AP, the solutions of (40) and (41) cost $\mathcal{O}((MN)^3)$ operations. The computational cost can be reduced to $\mathcal{O}(J^3 + N^3)$ for JLRS-SP and $\mathcal{O}(M^3 + N^3)$ for JLRS-SP when applying the Bartels-Stewart algorithm. On the other hand, both RPCA and LRR require an SVD step that has a computational complexity of $\mathcal{O}(JN^2)$ when using the inexact ALM optimization technique and a dictionary. Furthermore, LRR has an additional matrix inversion step, which is performed in the initialization stage. Therefore, the overall computational complexities of the JLRS-SP and JLRS-AP algorithms are $\mathcal{O}(t(J^3 + JN^2 + N^3))$ and $\mathcal{O}(t(M^3 + JN^2 + N^3))$, respectively, where t is the number of iterations. For RPCA, the overall computational complexity is $\mathcal{O}(tJN^2)$, whereas for LRR, it is $\mathcal{O}(tJN^2 + J^3)$. Though, the proposed algorithms are slightly more computationally expensive than RPCA and LRR, they have a more flexible model to determine the low-rank and sparse representations of the radar signal.

B. Regularization Parameters Tuning using Bayesian Optimization

The JLRS-AP and JLRS-SP models have three regularization parameters that control the amount of target information and clutter in the estimated matrix S. Setting the regularization parameters of the proposed method to large values discard the background clutter, at the expense of removing weak target reflections, and vice-versa. Therefore, the regularization parameters need to tune so that the sparse matrix S captures most of the target signature while maintaining low levels of clutter. In RPCA [31], the regularization parameter is defined as $1/\sqrt{\max(M,N)}$. In [42], the regularization parameter that links the mixed $\ell_{2,1}$-norm to the nuclear norm term is set to $3/(7\sqrt{\gamma \max(M,N)}$, where γ is a pre-defined constant. These two formulae may not be appropriate for the proposed method, due to the difference in the mathematical formulation of the JLRS model. A cross-validation grid search can be employed to determine the regularization parameters, but it is time consuming process when the searching boundary is large. Bayesian optimization, on the other hand, has been shown to obtain better results than grid search and random search [47], [48]. In conjunction with Gaussian process, Bayesian optimization has been used for tuning hyperparameters of machine learning methods, such as convolutional neural networks [49], support vector machines [50], and deep belief networks [51] as it is well-suited for global optimization problem where the objective function does not have an exact functional form and is computationally expensive to evaluate. Here, Bayesian optimization with Gaussian process is applied to determine the optimal regularization parameters for the JLRS-AP and JLRS-SP models.

Let $f(\boldsymbol{\lambda})$ denote the objective function that produces the quality score of the sparse matrix S (here, it is defined as the target to clutter ratio) obtained from the JLRS model with the set of regularization parameters $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3] \in \mathcal{Y}$. Bayesian optimization aims to find the regularization parameters $\boldsymbol{\lambda}$ that globally maximize f, i.e.,

$$\boldsymbol{\lambda}_* = \arg\max_{\boldsymbol{\lambda} \in \mathcal{Y}} f(\boldsymbol{\lambda}). \tag{45}$$

The Bayesian optimization algorithm requires a prior $p(f)$ over the function and an acquisition function $a : \mathcal{Y} \to \mathbb{R}^+$ to determine what point in $\mathcal{Y}$ should be evaluated next using a proxy optimization $\boldsymbol{\lambda}_* = \arg\max_{\boldsymbol{\lambda}} a(\boldsymbol{\lambda})$. To find the optimal regularization parameters, Bayesian optimization iterates the following three steps: (i) solve the proxy optimization $\boldsymbol{\lambda}_{t+1} = \arg\max_{\boldsymbol{\lambda} \in \mathcal{Y}} a(\boldsymbol{\lambda})$, (ii) evaluate the objective function $y_{t+1} \sim f(\boldsymbol{\lambda}_{t+1}) + \mathcal{N}(0, \sigma^2)$, which can be noisy and add the resulting data point $(\boldsymbol{\lambda}_{t+1}, y_{t+1})$ to the set of observations $\mathcal{D}_{t+1} = \{\boldsymbol{\lambda}_j, y_j\}_{j=1}^{t+1}$, and (iii) update $p(f|\mathcal{D}_{t+1})$ and $a(f|\mathcal{D}_{t+1})$. Gaussian process is a prominent choice for $p(f)$, due to its flexibility and tractability. It is specified by its mean function $m(\boldsymbol{\lambda})$ and covariance function $c(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j)$. Using the property of Gaussian distribution, the prior mean and covariance can be computed in a closed form. The prior mean function can be assumed to be zero in Gaussian process without any loss of generality; thereby, the Gaussian process can be fully defined by the covariance function. For hyperparameters, the ARD Matérn 5/2 kernel [52] is used as the covariance function and is given by

$$c_{M52}(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j) = \theta\big(1 + \sqrt{5} d(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j) + \frac{5}{3}(d(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j))^2\big) \exp\big(-\sqrt{5} d(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j)\big), \tag{46}$$

where $d(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j)$ is the Mahalanobis distance and θ is the characteristic length scale. The characteristic length scale defines how far apart the input $\boldsymbol{\lambda}$ can be for the response value to become uncorrelated.

Several acquisition functions have been proposed for Bayesian optimization, such as probability of improvement [53], expected improvement [54], and upper confidence bound [55]. Here, the expected improvement is used for acquisition function. Let us assume that the optimization problem is $\arg\max_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda})$ and the current best observation at iteration t is $\boldsymbol{\lambda}_* = \arg\max_{\boldsymbol{\lambda}_i \in \mathcal{Y}_{1:t}} f(\boldsymbol{\lambda})$. The improvement function is given by

$$J(\boldsymbol{\lambda}) = \max\big(0, f(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda}_*)\big). \tag{47}$$

The acquisition function is defined on the expected value of $J(\boldsymbol{\lambda})$ as

$$\arg\max_{\boldsymbol{\lambda}} \mathbb{E}\big(J(\boldsymbol{\lambda})|\mathcal{D}_t\big). \tag{48}$$

The closed form of $\mathbb{E}\big(J(\boldsymbol{\lambda})\big)$ can be written as [54]

$$\mathbb{E}\big(J(\boldsymbol{\lambda})\big)= \begin{cases} \big(m(\boldsymbol{\lambda})-f(\boldsymbol{\lambda}_*)\big)\eta(z)+s(\boldsymbol{\lambda})\kappa(z), & \text{if } s(\boldsymbol{\lambda})>0 \\ 0, & \text{if } s(\boldsymbol{\lambda})=0, \end{cases} \tag{49}$$

where $z=\big(m(\boldsymbol{\lambda})-f(\boldsymbol{\lambda}_*)\big)/s(\boldsymbol{\lambda})$, $s(\boldsymbol{\lambda})$ is the standard deviation function associated with the Gaussian process, $\eta(\cdot)$ and $\kappa(\cdot)$ are the cumulative distribution function and probability density function of a standard normal distribution, respectively.

C. Target Detection using Constant False Alarm Rate

The matrix S produced by the JLRS method is regarded as the target image, which comprises a certain number of non-zero columns. For target detection, a cell-averaged CFAR (CA-CFAR) detector is used to localize the non-zero columns containing the target responses. First, the matrix S is transformed into a saliency map M_s by computing two-dimensional (2D) discrete Fourier transform on local sliding windows:

$$M_s(i,j)=\max\big(\big|\text{FFT2}\{S(m,n),(m,n)\in W\}\big|\big) \quad i,m=1,\ldots,\ M,\ j,n=1,\ldots,N, \tag{50}$$

where FFT2 and W denote, respectively, 2D fast Fourier transform and the 2D local window centered at the location (m,n) in the matrix S. Then, the saliency map is converted into the spectral profile P, which is given by

$$P(j)=\frac{1}{M}\sum_{i=1}^{M} M_s(i,j), \quad j=1,\ldots,N. \tag{51}$$

The CA-CFAR detector is applied to the spectral profile to detect and localize the target signals. It is specified by the number of reference cells N_c surrounding the cell under test (CUT), which is used to estimate the clutter power, and the number of guard cells N_g on either side of the CUT. Let $\mathcal{X}_j$ denote the index set of the N_c reference cells surrounding the jth CUT. The jth column of the matrix S is detected as a target signal when the following condition is satisfied:

$$\frac{P(j)}{\sum_{i\in\mathcal{X}_j} P(i)} \geq T_0, \quad j=1,\ldots,N, \tag{52}$$

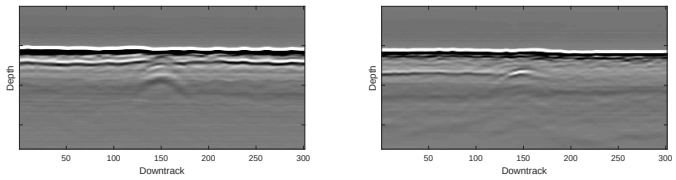
where T_0 is a predefined threshold.

IV. Experimental Results

In this section, the proposed JLRS-AP and JLRS-SP methods are evaluated on real GPR data for background clutter removal and target detection. First, the experimental setup is described, followed by performance analysis in terms of dictionary type. Then, the JLRS methods are compared with other existing methods for background clutter mitigation and target detection.

A. Experimental Setup

A NIITEK GPR array system is used to acquire GPR signals reflected from improvised explosive devices (IEDs) and landmines buried at different operation depths under the ground in a mild temperate terrain in Australia. These targets have different sizes and different amounts of metal content. They are grouped into two categories: small targets of size less than 50 mm and large targets of size greater than 50 mm. The GPR signals are collected along the down-track direction and arranged into B-scans, see Fig. 1. Each B-scan has a size of 180×301, i.e., 180 depth bins and 301 A-scans at an interspace of 0.05 m. Moreover, each B-scan has a single target positioned at the 150th column of the B-scan. Each A-scan is re-scaled to the range [0, 1] and then centered by subtracting the mean. A database of 574 down-track B-scans collected from 27 different types of IEDs and landmines is used to evaluate the proposed JLRS method for clutter removal and target detection. To tune the regularization parameters and compute the noise covariance matrix of the JLRS models, a validation set comprising 135 B-scans is generated using five B-scans per target type. The remaining 439 B-scans are reserved for the test set. The test set contains 328 large targets and 111 small targets.



(a) (b)

Fig. 1. B-scans containing GPR traces collected along the down-track direction in which the target is buried (a) close to ground surface and (b) at a certain operational depth.

The improvement factor (IF) in terms of the target-to-clutter ratio (TCR) is used to measure the quality of the target image produced by the background clutter removal method and is computed as

$$\text{IF}=\frac{\text{TCR}_a}{\text{TCR}_b}, \tag{53}$$

where TCR_a and TCR_b are, respectively, the target-to-clutter ratios of the B-scan after and before background clutter removal. The TCR of a B-scan is calculated as

$$\text{TCR}=\frac{N_c\sum_{(i,j)\in A_t}|S(i,j)|^2}{N_t\sum_{(i,j)\in A_c}|S(i,j)|^2}, \tag{54}$$

where A_t is the selected target region, A_c is the clutter region defined as the entire image excluding the target region, N_c and N_t are, respectively, the number of pixels in the clutter and target regions. For a set of K B-scans, the average improvement factor IF_{av} in decibel (dB) is computed as

$$\text{IF}_{av}=10\log_{10}\Big(\frac{1}{K}\sum_{i=1}^{K}\text{IF}(i)\Big). \tag{55}$$

TABLE I
AVERAGE IMPROVEMENT FACTOR (IN DB) OF JLRS-AP AND JLRS-SP MODELS, USING DIFFERENT COMBINATIONS OF THE LOW-RANK (Φ AND Θ) AND SPARSITY (Ψ AND Ω) DICTIONARIES.

Dictionary	**JLRS-AP**			**Dictionary**	**JLRS-SP**		
Φ \ Ψ	DFT	DT-CWT	WPD	Θ \ Ω	DFT	DT-CWT	WPD
DFT	10.82	9.28	19.11	DFT	10.82	9.28	11.11
DT-CWT	9.68	6.78	**19.30**	DT-CWT	9.68	6.78	10.41
WPD	10.23	9.23	15.24	WPD	10.23	9.23	**13.14**

Furthermore, the receiver operating characteristic (ROC) curve of the CFAR detector in conjunction with a background clutter removal method is computed to evaluate the detection accuracy. The ROC curve is a graphical representation of the accuracy of a detector and is created by plotting the probability of detection against the probability of false alarm. The probability of detection (PD) is calculated as the ratio of the number of positive detections to the total number of targets. A positive detection is declared when the detected peak of the spectral profile is within the predefined target region. The probability of false alarm (PFA), on the other hand, is computed as the ratio of the number of detected signals outside the target region to the total number of non-target signals, where radar signals outside the target region are considered as non-target signals.

B. Effect of Dictionary Type on the Performance of JLRS-AP and JLRS-SP

Different types of dictionaries can be employed to determine the low-rank and sparse matrices. Here, three types of signal transforms are investigated for generating the dictionaries: discrete Fourier transform (DFT), dual-tree complex wavelet transform (DT-CWT), and wavelet packet decomposition (WPD). The rationale for using these signal transforms is as follows. DFT and DT-CWT produce signal coefficients whose magnitudes are tolerant to shift-variations. WPD can be used to represent a signal by a small number of non-zero coefficients. The DFT dictionary is square and comprises orthonormal atoms, whereas the WPD dictionary is an overcomplete set of wavelet atoms. Based on the preliminary analysis, the discrete approximation of Meyer wavelet was found to achieve the highest improvement factor. Therefore, it is used to produce a WPD dictionary using four levels of decomposition. The same number of levels of decomposition is also used to generate a square DT-CWT dictionary.

All three dictionaries such as DFT, DT-CWT, and WPD are then evaluated to determine the appropriate dictionaries for estimating the target signal representation. For each dictionary combination, the Bayesian optimization technique is used to tune the regularization parameters. The search intervals of the regularization parameters are set as follows: $0.001 \leq \lambda_1 \leq 0.9$, $0.001 \leq \lambda_2 \leq 0.9$, and $0.00001 \leq \lambda_3 \leq 0.1$. The optimization technique is stopped when either the relative difference between two consecutive low-rank matrices is below the predefined threshold δ, i.e., $\left\|L^{k+1} - L^k\right\|_F / \|Y\|_F \leq 0.01$ or the number of iterations is equal to 100. The proposed JLRS methods with the optimal regularization parameters are then evaluated on the validation set. Table I presents the average improvement factor (IF_{av}) of the JLRS-AP and JLRS-SP models for different combinations of the low-rank and sparsity dictionaries. Both models achieve the same IF_{av} when using either DFT or DT-CWT dictionary for sparse representation. This is not surprising because the two models are equivalent when the dictionary Ω is invertible ($\Omega = \Psi^{-1}$), which is the case for the DFT and DT-CWT dictionaries. However, the two models are not equivalent if Ψ is an overcomplete dictionary (e.g., WPD) and $\Omega = \Psi^{\dagger}$. The analysis model achieves better IF_{av} than the synthesis model. This is because JLRS-AP emphasizes the zero coefficients of the sparse representation and exploits the zero-crossing of wavelet transform, thereby requiring fewer wavelet packet atoms to represent the target signal, compared to JLRS-SP. The best dictionary combination for JLRS-AP is DT-CWT dictionary for low-rank and WPD dictionary for sparsity. For JLRS-SP, the best combination is to use the WPD dictionary for both low-rank and sparsity.

C. Comparison of Different Background Clutter Mitigation Methods

For comparison purposes, five baseline methods were implemented: SVD, PCA, ICA, RPCA, and LRR. They were all tested on the same test set. The rationale for comparing these baseline methods is that they adopt similar concept for background clutter removal, i.e., capturing the background clutter in a subspace. In the SVD and PCA methods, the number of components spanning the clutter subspace was manually varied from 1 to 10. In the ICA method [25], SVD was firstly used to pre-whiten the B-scan. Then, the FASTICA algorithm [56] was applied to determine the mixing matrix and the independent components. The normalized kurtosis was employed to identify the independent components spanning the target subspace. Contrary to [29], which performed B-scan alignment and truncation of the ground surface clutter before low-rank and sparse signal decomposition, LRR and RPCA were applied directly to the B-scan. Moreover, different dictionaries, namely DFT, DT-CWT, and WPD were used to determine the low-rank and sparse representations. Optimization methods based on inexact ALM were implemented to solve the RPCA and LRR problems.

Figure 4 depicts the IF_{av} values of the SVD and PCA methods as a function of the number of components spanning the clutter subspace. Increasing the number of singular vectors and eigen-vectors improves the IF_{av} values of the

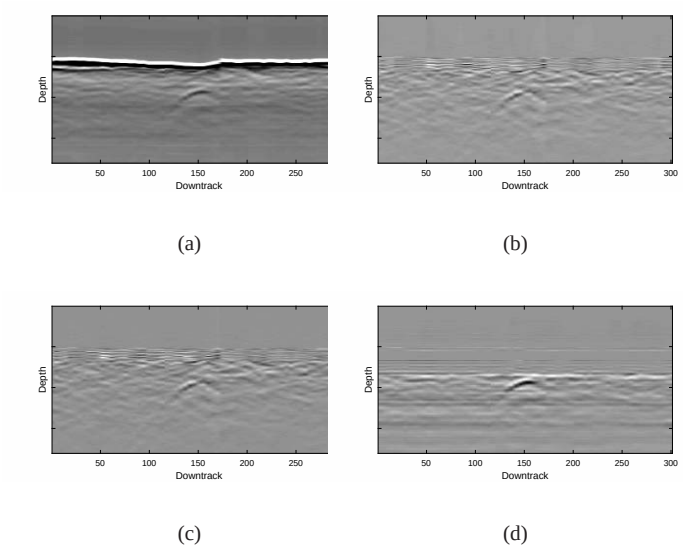


Fig. 2. Target B-scans obtained from the (a) input B-scan using the standard subspace methods: (b) SVD, (c) PCA, and (d) ICA.

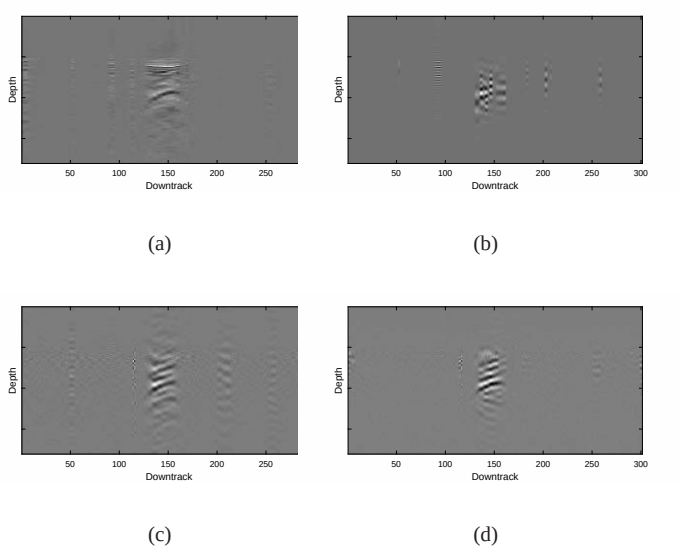


Fig. 3. Target B-scans produced by the JLRS methods: (a) LRR, (b) RPCA, (c) JLRS-SP and (d) JLRS-AP.

SVD and PCA methods. For PCA, the clutter subspace is spanned by 7 eigen-vectors, whereas the subspace obtained using SVD is spanned by 8 singular vectors. Table II lists their IF$_{av}$ values and those of the ICA and JLRS methods. Both SVD and PCA achieve similar IF$_{av}$ values of 6.13 dB and 6.10 dB, respectively. ICA, on the other hand, obtains higher IF$_{av}$ than SVD and PCA. RPCA and LRR outperform the standard subspace methods since they jointly estimate the low-rank and sparse matrices from the B-scan. Converting the GPR signals into another domain before low-rank and sparse signal decomposition gives better improvement factor than applying RPCA directly to the B-scan. Similar observations were made by Masarik *et al.* [29], where the GPR signals were transformed to the frequency domain. Among the three types of analysis dictionaries, WPD produces the best IF$_{av}$ of 18.03 dB for RPCA. LRR achieves an IF$_{av}$ of 6.12 dB when the input B-scan is used as the synthesis dictionary. However, using DT-CWT to form the synthesis dictionary for LRR improves the IF$_{av}$ to 14.87 dB. Other dictionaries such as WPD and DFT produce slightly lower IF$_{av}$ values than DT-CWT. Among the six background clutter removal techniques, the proposed JLRS method achieves the highest IF$_{av}$. JLRS-AP and JLRS-SP obtain IF$_{av}$ values of 19.31 dB and 13.41 dB, respectively.

Figure 2 shows the target B-scans produced by the standard subspace methods when applying to the input B-scan depicted in Fig. 2(a). In all three target B-scans illustrated in Figs. 2(b) to (d), the strong ground reflections have been removed. However, the background noise and other scatterings are still present in the target B-scans. The target reflections in the B-scans obtained from ICA are stronger than those in the B-scans produced by SVD and PCA as well as the background clutter. Reducing the number of dominant components spanning the clutter subspace strengthens the target reflections, at the expense of keeping more of the background clutter and noise. Figure 3 depicts the output B-scans of the JLRS

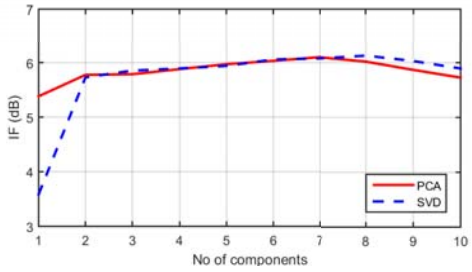


Fig. 4. The average improvement factor (IF$_{av}$) of the SVD and PCA methods as a function of the number of dominant components spanning the background clutter subspace.

methods. Figures 3(a) and (b) show the B-scans obtained using LRR and RPCA, whereas Figs. 3(c) and (d) present the B-scans produced by the proposed method. JLRS-AP and JLRS-SP, which employ both ℓ_1-norm and mixed $\ell_{2,1}$-norm sparsity constraints, generate much clearer target B-scans than the standard subspace methods. They not only remove the background clutter, but also preserve the target signature. Comparing the B-scans in Figs. 3(a) and (b) with those depicted in Figs. 3(c) and (d), we show that the JLRS-AP model is more effective than the RPCA and LRR models for background clutter removal. Since JLRS-AP is superior to JLRS-SP, it is used for target detection in the following experiments.

D. Detection Performance

For background clutter removal, the proposed JLRS-AP method outperforms the other baseline methods in terms of target-to-clutter ratio. In this experiment, it is combined with a CA-CFAR detector for target detection. The parameters of the detector are set as follows. The size of the sliding window W is 5×5, the number of reference cells N_c is 100, and the number of guards cell N_g is 40. The diameters of the

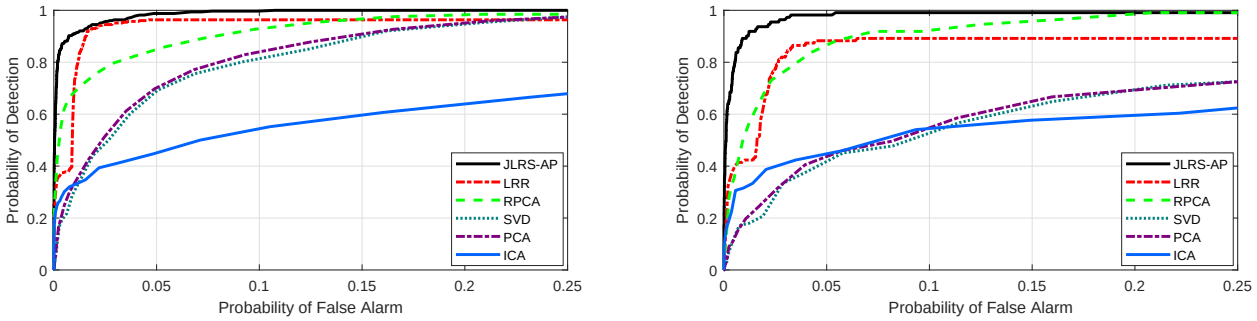


Fig. 5. ROC curves of the CFAR detector in conjunction with different background clutter removal methods evaluated on (a) large targets and (b) small targets.

TABLE II
AVERAGE IMPROVEMENT FACTOR OF DIFFERENT BACKGROUND CLUTTER MITIGATION METHODS EVALUATED ON THE TESTSET.

Background clutter mitigation method	IF_{av} (dB)
Proposed JLRS-AP method	**19.31**
Proposed JLRS-SP method	13.41
RPCA method [31]	8.47
RPCA method (Complex wavelet domain)	13.81
RPCA method (Wavelet domain)	18.03
RPCA method (Frequency domain)	14.18
LRR method [32]	6.12
LRR method (DT-CWT dictionary)	14.87
LRR method (WPD dictionary)	14.04
LRR method (DFT dictionary)	14.12
SVD	6.13
PCA	6.10
ICA	7.57

pre-defined regions for positive detection of large and small targets are 2.05 and 1.05 m, respectively. The threshold T_0 is varied to generate different values of probability of detection (PD) and probability of false alarm (PFA) for plotting the ROC curve. The confidence interval of the PD is computed based on the assumption that the PDs are binomially distributed, as described in [57].

Figure 5 shows the ROC curves of the CA-CFAR detector applied to the target B-scans generated by different background clutter removal techniques. Figure 5(a) depicts the ROC curves obtained from B-scans with large targets and Fig. 5(b) shows those from B-scans with small targets. Tested on large buried targets at fixed PFA of 0.05, JLRS-AP achieves a PD of 0.988, followed by LRR with a PD of 0.963 and RPCA with a PD of 0.859. The subspace methods achieve a much lower PD, see Table III. The PDs for small targets are 0.982 for JLRS-AP, 0.883 for LRR, 0.882 for RPCA, 0.459 for PCA, 0.450 for SVD, and 0.459 for ICA.

The area under the ROC curve (AUC) is also computed to measure the detection accuracy of the proposed method. Table IV presents the AUCs and the 95% confidence intervals obtained from the different background clutter removal methods in conjunction with the CA-FAR detector. The AUC is computed using the trapezoidal integration technique and the confidence interval is obtained using the method developed by Hanley and McNeil [58]. JLRS-AP has the highest detection accuracy, compared to other methods. JLRS-AP achieves AUCs of 0.980 and 0.979 for large and small targets, respectively, followed by RPCA and LRR. Among the subspace methods, ICA obtains the lowest AUC.

TABLE III
PROBABILITY OF DETECTION (PD) OF THE CA-FAR DETECTOR AT FIXED FALSE ALARM RATE OF 0.05, USING DIFFERENT BACKGROUND CLUTTER REMOVAL METHODS.

Method	Large target		Small target	
	PD	95% CI	PD	95% CI
JLRS-AP	**0.988**	[0.969, 0.997]	**0.982**	[0.936, 0.998]
LRR + CWT	0.963	[0.937, 0.981]	0.883	[0.808, 0.936]
RPCA + WPD	0.859	[0.817, 0.895]	0.882	[0.807, 0.935]
PCA	0.698	[0.645, 0.747]	0.459	[0.364, 0.556]
SVD	0.692	[0.639, 0.742]	0.450	[0.355, 0.547]
ICA	0.448	[0.393, 0.504]	0.459	[0.364, 0.556]

TABLE IV
AREA UNDER THE ROC CURVE AND CONFIDENCE INTERVALS OF THE CA-FAR DETECTOR USING DIFFERENT BACKGROUND CLUTTER REMOVAL METHODS.

Method	Large target		Small target	
	AUC	95% CI	AUC	95% CI
JLRS-AP	**0.980**	[0.970,0.991]	**0.979**	[0.961, 0.998]
LRR + CWT	0.941	[0.922, 0.960]	0.873	[0.829, 0.918]
RPCA + WPD	0.960	[0.945, 0.975]	0.959	[0.934, 0.985]
PCA	0.934	[0.914, 0.954]	0.817	[0.759, 0.875]
SVD	0.930	[0.909, 0.951]	0.808	[0.756, 0.861]
ICA	0.794	[0.758, 0.829]	0.758	[0.701, 0.815]

V. Conclusion

In this paper, a joint low-rank and sparse signal decomposition method was proposed for background clutter removal and target detection. The proposed method is based on the observations that the background clutter resides in a low-rank subspace and the target signals, which tend to be clustered in few non-zero columns, form a sparse matrix. Thus, the background clutter mitigation and target detection is formulated as a joint low-rank and sparsity constrained optimization problem, which is solved using the inexact augmented Lagrangian multiplier method. Contrary to the mathematical models of the RPCA and LRR techniques, the proposed method adopts the analysis or synthesis prior to estimating the low-rank and sparse representations. The use of these priors avoids the conversion of the B-scan into another domain before signal decomposition and provides the flexibility to use different dictionaries for low-rank and sparse signal decomposition. Furthermore, the proposed JLRS method takes into account the noise of the B-scan. Experiments were conducted using real GPR data collected from buried landmines and improvised explosive devices. Experimental results showed that the proposed method using analysis prior achieved better results than using synthesis prior and outperformed the existing standard subspace methods and joint low-rank and sparse methods, such as RPCA and LRR techniques.

Acknowledgment

This work was supported by grants from the Defence Science and Technology Group (DSTG) and the Australian Research Council (ARC).

VI. Appendix

The last four terms on the right hand side of the augmented Lagrangian function given in (14) can be simplified as follows:

$$
\begin{aligned}
&\langle B, Z-F\rangle+\langle C, X-G\rangle+\frac{\mu}{2}\left[\left\|Z-F\right\|_F^2+\left\|X-G\right\|_F^2\right]\\
&=\mathrm{tr}[B^T(Z-F)]+\mathrm{tr}[C^T(X-G)]+\\
&\qquad\qquad\frac{\mu}{2}\left[\left\|Z-F\right\|_F^2+\left\|X-G\right\|_F^2\right]\\
&=\mathrm{tr}[B^T(Z-F)]+\mathrm{tr}[C^T(X-G)]+\\
&\qquad\frac{\mu}{2}\mathrm{tr}[(Z-F)^T(Z-F)+(X-G)^T(X-G)]\\
&=\frac{\mu}{2}\mathrm{tr}[\frac{2}{\mu}(B^T(Z-F))+\frac{2}{\mu}(C^T(X-G))+\\
&\qquad\qquad(Z-F)^T(Z-F)+(X-G)^T(X-G)]\\
&=\frac{\mu}{2}\mathrm{tr}[\frac{2}{\mu}(B^T(Z-F))+\frac{2}{\mu}(C^T(X-G))+\\
&\qquad(Z-F)^T(Z-F)+(X-G)^T(X-G)+\\
&\qquad\qquad\frac{B^TB}{\mu^2}-\frac{B^TB}{\mu^2}+\frac{C^TC}{\mu^2}-\frac{C^TC}{\mu^2}\\
&=\frac{\mu}{2}\left\|Z-F+\frac{B}{\mu}\right\|_F^2+\frac{\mu}{2}\left\|X-G+\frac{C}{\mu}\right\|_F^2-\\
&\qquad\qquad\frac{1}{2\mu}\left[\left\|B\right\|_F^2+\left\|C\right\|_F^2\right]
\end{aligned}
$$

The main steps of the ALM-based optimization techniques for solving the JLRS-SP and JLRS-AP models are presented in the following algorithms:

Algorithm 1 - JLRS-SP optimization algorithm

Input: Y, Z^0, X^0, Θ, Ω, $G^0=B^0=C^0=0$, λ_1, λ_2, λ_3, $\mu_0=0.25*\sum_{i=1}^{M}\sum_{j=1}^{N}|Y(i,j)|/MN$, $\mu_{\max}=10^{10}$, $\rho=1.01$, $\delta=0.01$, $N_{iter}=30$.

Output: L, S

for $k=1$ to N_{iter} **do**
- $\mathrm{SVD}\Big(Z^k+B^k/\mu_k\Big)=UDV^T$;
- $F^{k+1}=U\mathcal{T}\big(D,1/\mu_k\big)V^T$;
- $\boldsymbol{d}^{k+1}=\mathrm{vec}(\lambda_3\Theta^T(Y-\Omega X^k)+\mu^k(F^{k+1}-B^k/\mu^k)\Sigma)$
- $Z^{k+1}=\mathrm{mat}\Big([I\otimes(\lambda_3\Theta^T\Theta)+\mu^k\Sigma^T\otimes I]^{-1}\boldsymbol{d}^{k+1}\Big)$;
- $\widehat{\boldsymbol{d}}^{k+1}=\mathrm{vec}(\lambda_3\Omega^T(Y-\Theta Z^{k+1})+\mu^k(G^k-C^k/\mu^k)\Sigma)$
- $X^{k+1}=\mathrm{mat}\Big([I\otimes(\lambda_3\Omega^T\Omega)+\mu^k\Sigma^T\otimes I]^{-1}\widehat{\boldsymbol{d}}^{k+1}\Big)$;
- $\boldsymbol{g}_i^{k+1}=\mathcal{G}\big(\boldsymbol{x}_i^{k+1}+\boldsymbol{c}_i^k/\mu_k,\gamma_1,\gamma_2\big)$;
- $B^{k+1}=B^k+\mu^k\big(F^{k+1}-Z^{k+1}\big)$;
- $C^{k+1}=C^k+\mu^k\big(X^{k+1}-G^{k+1}\big)$;
- $\mu^{k+1}=\min\big(\rho\mu^k,\mu_{\max}\big)$;
- **if** $\left\|\Theta(Z^{k+1}-Z^k)\right\|_F/\left\|Y\right\|_F\le\delta$ **then**
  - Break;
- **end if**

end for

Algorithm 2 - JLRS-AP optimization algorithm

Input: Y, L^0, S^0, R^0, Φ, Ψ, $B^0=C^0=0$, λ_1, λ_2, λ_3, $\mu_0=0.25*\sum_{i=1}^{M}\sum_{j=1}^{N}|Y(i,j)|/MN$, $\mu_{\max}=10^{10}$, $\rho=1.01$, $\delta=0.01$, $N_{iter}=30$.

Output: L, S

for $k=1$ to N_{iter} **do**
- $\mathrm{SVD}\Big(\Phi L^k+B^k/\mu_k\Big)=UDV^T$;
- $Q^{k+1}=U\mathcal{T}\big(D,1/\mu_k\big)V^T$;
- $\boldsymbol{d}^{k+1}=\mathrm{vec}(\lambda_3(Y-S^k)\Sigma^{-1}+\mu^k\Phi^T(Q^{k+1}-B^k/\mu^k))$
- $L^{k+1}=\mathrm{mat}\Big([I\otimes(\mu^k\Phi^T\Phi)+(\lambda_3\Sigma^{-1})^T\otimes I]^{-1}\boldsymbol{d}^{k+1}\Big)$;
- $\widehat{\boldsymbol{d}}^{k+1}=\mathrm{vec}(\lambda_3(Y-L^{k+1})\Sigma^{-1}+\mu^k\Psi^T(R^k-C^k/\mu^k))$
- $S^{k+1}=\mathrm{mat}\Big([I\otimes(\mu^k\Psi^T\Psi)+(\lambda_3\Sigma^{-1})^T\otimes I]^{-1}\widehat{\boldsymbol{d}}^{k+1}\Big)$;
- $\boldsymbol{r}_i^{k+1}=\mathcal{G}\big(\Psi\boldsymbol{s}_i^{k+1}+\boldsymbol{c}_i^k/\mu^k,\gamma_1,\gamma_2\big)$;
- $B^{k+1}=B^k+\mu^k\big(\Phi L^{k+1}-Q^{k+1}\big)$;
- $C^{k+1}=C^k+\mu^k\big(\Psi S^{k+1}-R^{k+1}\big)$;
- $\mu^{k+1}=\min\big(\rho\mu^k,\mu_{\max}\big)$;
- **if** $\left\|L^{k+1}-L^k\right\|_F/\left\|Y\right\|_F\le\delta$ **then**
  - Break;
- **end if**

end for

References

[1] A. M. Zoubir, I. J. Chant, C. L. Brown, B. Barkat, and C. Abeynayake, "Signal processing techniques for landmine detection using impulse ground penetrating radar," *IEEE Sensors Journal*, vol. 2, no. 1, pp. 41–51, 2002.

[2] Z. W. Wang, M. Zhou, G. G. Slabaugh, J. Zhai, and T. Fang, "Automatic detection of bridge deck condition from ground penetrating radar images," *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 3, pp. 633–640, 2010.

[3] J. S. Wei and M. Hashim, "Accuracy of data acquisition approaches with ground penetrating radar for subsurface utility mapping," in *IEEE International RF and Microwave Conference*, Seremban, Negeri Sembilan, Malaysia, 12-14 Jun., 2011, pp. 40–44.

[4] 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 Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3961–3972, 2011.

[5] C. Trela, T. Kind, and M. Schubert, "Detection of air voids in concrete by radar in transmission mode," in *Proc. 8th International Workshop on Advanced Ground Penetrating Radar*, Florence, Italy, 7-10 Jul., 2015, pp. 1–4.

[6] D. Comite, F. Ahmad, M. G. Amin, and T. Dogaru, "Detection of low-signature targets in rough surface terrain for forward-looking ground penetrating radar imaging," in *IEEE 49th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 8-11 Nov., 2015, pp. 80-84.

[7] D. Comite, F. Ahmad, T. Dogaru, and M. G. Amin, "Coherence factor for rough surface clutter mitigation in forward-looking GPR," in *IEEE Radar Conference*, Seattle, WA, USA, 8-12 May, 2017, pp. 1803-1506.

[8] S.-E. Hamran, D. T. Gjessing, J. Hjelmstad, and E. Aarholt, "Ground penetrating synthetic pulse radar: dynamic range and modes of operation," *Journal of Applied Geophysics*, vol. 33, no. 1-3, pp. 7–14, 1995.

[9] R. Solimene, A. Cuccaro, A. Dell'Aversano, I. Catapano, and F. Soldovieri, "Ground clutter removal in GPR surveys," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 3, pp. 792–798, 2014.

[10] B. M. Duston and D. A. Lang, "Statistical processing of ground penetrating radar signals for mine detection," in *Proc. SPIE, Detection and Remediation Technologies for Mines and Minelike Targets VI,*, vol. 4394, Orlando, FL, USA, 16 Apr., 2001, pp. 494–502.

[11] D. Potin, E. Duflos, and P. Vanheeghe, "Landmines ground-penetrating radar signal enhancement by digital filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2393–2406, 2006.

[12] T.-J. Li, L.-J. Kong, and Z.-O. Zhou, "Symmetry filtering method for GPR clutter reduction," in *International Conference on Microwave and Millimeter Wave Technology*, Nanjing, China, 21-24 Apr., 2008, pp. 1515–1517.

[13] F. Abujarad, G. Nadim, and A. Omar, "Wavelet packets for GPR detection of non-metallic anti-personnel land mines based on higher-order-statistic," in *Proc. of the 3rd International Workshop on Advanced Ground Penetrating Radar*, Delft, Netherlands, 2-4 May, 2005, pp. 21–25.

[14] G. Terrasse, J.-M. Nicolas, E. Trouve, and m. Drouet, "Application of the curvelet transform for pipe detection in GPR images," in *IEEE International Geoscience and Remote Sensing Symposium*, Milan, Italy, 26-31 Jul., 2015, pp. 4308–4311.

[15] H. Brunzell, "Detection of shallowly buried objects using impulse radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 875–886, 1999.

[16] A. v. d. Merwe and I. J. Gupta, "A novel signal processing technique for clutter reduction in GPR measurements of small, shallow land mines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 6, pp. 2627–2637, 2000.

[17] T. C. T. Chan, H. C. So, and K. Ho, "Generalized two-sided linear prediction approach for land mine detection," *Signal Processing*, vol. 88, no. 4, pp. 1053–1060, 2008.

[18] B. Cagnoli and T. J. Ulrych, "Singular value decomposition and wavy reflections in ground-penetrating radar images of base surge deposits," *Journal of Applied Geophysics*, vol. 48, no. 3, pp. 175–182, 2001.

[19] F. Abujarad, A. Jostingmeier, and A. S. Omar, "Clutter removal for landmine using different signal processing techniques," in *Proc. Tenth International Conference on Ground Penetrating Radar*, Delft, Netherlands, 21-24 Jun., 2004, pp. 697–700.

[20] M. M. Riaz and A. Ghafoor, "Information theoretic criterion based clutter reduction for ground penetrating radar," *Progress In Electromagnetics Research B*, vol. 45, pp. 147–164, 2012.

[21] M. E. Yavuz, A. E. Fouda, and F. L. Teixeira, "GPR signal enhancement using sliding-window space-frequency matrices," *Progress In Electromagnetics Research*, vol. 145, pp. 1–10, 2014.

[22] B. Karlsen, J. Larsen, H. B. D. Sorensen, and K. B. Jakobsen, "Comparison of PCA and ICA based clutter reduction in GPR systems for anti-personal landmine detection," in *Proc. 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Singapore, 6-8 Aug., 2001, pp. 146–149.

[23] J. Liu, R. Wu, T. Li, and B. Zhang, "Novel ground bounce removal algorithms based on non-homogeneous detector," in *International Conference on Radar*, Shanghai, China, 16-19 Oct., 2006, pp. 1–5.

[24] Q. Q. Lu, J. X. Pu, X. H. Wang, and Z. H. Liu, "A clutter suppression algorithm for GPR data based on PCA combining with gradient magnitude," *Applied Mechanics and Materials*, vol. 644-650, pp. 1662–1667, 2014.

[25] B. Karlsen, H. B. D. Sorensen, J. Larsen, and K. B. Jakobsen, "GPR detection of buried symmetrically shaped minelike objects using selective independent component analysis," in *Proc. SPIE, Detection and Remediation Technologies for Mines and Minelike Targets VIII*, vol. 5089, Orlando, FL, USA, 12-25 Apr., 2003, pp. 375–386.

[26] A. Zhao, Y. Jiang, and W. Wang, "Exploring independent component analysis for GPR signal processing," *Progress In Electromagnetics Research Symposium*, pp. 750–753, Hangzhou, China, 22-26 Aug., 2005.

[27] Q. Gao, T. Li, and R. Wu, "A novel KICA method for ground bounce removal with GPR," in *International Conference on Radar*, Shanghai, China, 16-19 Oct., 2006, pp. 1–4.

[28] W. Chen, W. Wang, J. Gao, J. Xu, and W. Wang, "GPR clutter noise separation by statistical independency promotion," in *Proc. 14th International Conference on Ground Penetrating Radar*, Shanghai, China, 4-8 Jun., 2012, pp. 367–370.

[29] M. P. Masarik, J. Burns, B. T. Thelen, J. Kelly, and T. C. Havens, "GPR anomaly detection with robust principal component analysis," in *Proc. SPIE, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX*, vol. 9454, Baltimore, Maryland, USA, 20-23 Apr., 2015, pp. 945 414 1–945 414 11.

[30] A. Pinar, T. C. Havens, J. Rice, M. Masarik, J. Burns, and B. Thelen, "A comparison of robust principal component analysis techniques for buried object detection in downward looking GPR sensor data," in *Proc. SPIE, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXI*, vol. 9823, Baltimore, Maryland, USA, 18-21 Apr., 2016, pp. 98 230T 1– 98 230T 16.

[31] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, Article no 11, pp. 1–37, 2011.

[32] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2012.

[33] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[34] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," University of Illinois, Technical Report No. UILU-ENG-09-2215, 2009.

[35] J. F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[36] A. Ganesh, Z. Lin, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast algorithms for recovering a corrupted low-rank matrix," in *International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Aruba, Dutch Antilles, The Netherlands, 13-16 Dec., 2009, pp. 213–216.

[37] B. Wohlberg, R. Chartrand, and J. Theiler, "Local principal component pursuit for nonlinear data sets," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Kyoto, Japan, 25-30 Mar., 2012, pp. 3925–3928.

[38] J. F. Yang and Y. Zhang, "Alternating direction algorithms for L1-problems in compressive sensing," *SIAM Journal on Scientific Computing*, vol. 33, no. 1, pp. 250–278, 2011.

[39] Z. Yang, C. Zhang, J. Deng, and W. Lu, "Orthonormal expansion L1-minimization algorithms for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6285–6290, 2011.

[40] H. E. Guven, A. Gungor, and M. Cetin, "An augmented Lagrangian method for image reconstruction with multiple features," *IEEE International Conference on Image Processing*, pp. 4175–4179, Quebec City, Canada, 27-30 Sept., 2015.

[41] M. Afonso and J. M. Sanches, "Image reconstruction under multiplicative speckle noise using total variation," *Neurocomputing*, vol. 150, no. Part A, pp. 200–213, 2015.

[42] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3047–3064, 2012.

[43] D. Fernandez and M. V. Solodov, "Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition," *SIAM Journal on Optimization,*, vol. 22, no. 2, pp. 384–407, 2012.

[44] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[45] A. Majumdar and R. K. Ward, "Some empirical advances in matrix completion," *Signal Processing*, vol. 91, no. 5, pp. 1334–1338, 2011.

[46] F. H. C. Tivive, A. Bouzerdoum, and M. G. Amin, "A subspace projection approach for wall clutter mitigation in through-the-wall radar imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2108–2122, 2015.

[47] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kegl, "Algorithms for hyper-parameter optimization," in *Proc. of the 24th International Conference on Neural Information Processing Systems*, Granada, Spain, 12-15 Dec., 2011, pp. 2546–2554.
[48] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. of the 25th International Conference on Neural Information Processing Systems*, vol. 2, Lake Tahoe, Nevada, USA, 3-6 Dec., 2012, pp. 2951–2959.
[49] J. Snoek, O. Rippel, k. Swersky, R. Kiros, N. Satish, N. Sundaram, M. M. A. Patwary, P. Prabhat, and R. P. Adams, "Scalable Bayesian optimization using deep neural networks," in *Proc. of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, Lille, France, 6-11 Jul., 2015, pp. 2171–2180.
[50] W. M. Czarnecki, S. Podlewska, and A. J. Bojarski, "Robust optimization of SVM hyperparameters in the classification of bioactive compounds," *Journal of Cheminformatics*, vol. 7, no. 38, pp. 1–15, 2015.
[51] C. Yao, D. Cai, J. Bu, and G. Chen, "Pre-training the deep generative models with adaptive hyperparameter optimization," *Neurocomputing*, vol. 247, no. 19, pp. 144–155, 2017.
[52] B. Matern, *Spatial variation*, ser. Meddelanden frans Statens Skogsforskningsinstitut, vol. 49, 1960.
[53] H. J. Kushner, "A new method for locating the maximum point of an arbitrary multipeak curve in the presence of noise," *Journal of Basic Engineering*, vol. 86, no. 1, pp. 97–106, 1964.
[54] J. Mockus, V. Tiesis, and A. Zilinskas, "The application of Bayesian methods for seeking the extremum," in *Towards Global Optimisation*, L.C. W Dixon and G. P. Szego, Eds., North-Holland Pub. Co., 1978, pp. 117–130.
[55] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: no regret and experimental design," in *Proc. of the 27th International Conference on International Conference on Machine Learning*, Haifa, Israel, 21-24 Jun., 2010, pp. 1015–1022.
[56] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.
[57] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 251–255, 2008.
[58] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

Fok Hing Chi Tivive received the BE(Hons) in Telecommunications from Edith Cowan University and the Ph.D degree in computer engineering from the University of Wollongong, Australia, in 2001 and 2006, respectively. Since 2006, he has been with the School of Electrical Computer and Telecommunications Engineering, University of Wollongong, as a Postdoctoral Research Fellow. His research interests include machine learning, pattern recognition, image processing, ground penetrating radar, and through-the-wall radar imaging.

Abdesselam Bouzerdoum (M'89-SM'03 IEEE) graduated with MSEE and Ph.D. degrees from the University of Washington, Seattle, USA. In 1991, he joined Adelaide University, South Australian, and from 1998 to 2004 he was an Associate Professor with Edith Cowan University, Perth, Western Australia. In September 2004, he was appointed Professor of Computer Engineering and Head of School of Electrical, Computer & Telecommunications Engineering at the University of Wollongong. From 2007 to 2013, he served as Associate Dean (Research), Faculty of Informatics. Since March 2017, he has also been a Professor with Hamad Bin Khalifa University (HBKU), Qatar. Dr. Bouzerdoum held several Visiting Professor Appointments at Institut Galile, Universit Paris-13, LAAS/CNRS, Toulouse, France, Villanova University, USA, and the Hong Kong University of Science and Technology. From 2009 to 2011, he was a member of the ARC College of Experts and Deputy Chair of the EMI panel (2010-2011).

Dr. Bouzerdoum is the recipient of the *Eureka Prize for Outstanding Science in Support of Defence or National Security* (2011), the *Chester Sall Award of IEEE Trans. Consumer Electronics* (2005), and a *Distinguished Researcher Award (Chercheur de Haut Niveau) from the French Ministry* (2001). He served as Associate Editor for 5 International journals, including IEEE TRANS. IMAGE PROCESSEING, IEEE TRANS. SYSTEMS, MAN, AND CYBERNETICS (1999-2006). He has published over 340 technical articles and graduated many Ph.D. and Research Masters students. His research interest include radar imaging and signal processing, image processing, vision, machine learning, and pattern recognition.

Canicious Abeynayake received the Ph.D. degree in physics from Peoples' Friendship University, Moscow, Russia, in 1989 and the Ph.D. degree in signal and information processing from the University of South Australia, Adelaide, Australia, in 2002.

From 1989 to 1995 he was a senior lecturer with the University of Rzeszow, Poland. Before joining the Defence Science and Technology (DST) Group, Edinburgh in Australia, he held academic positions in Australia. He has been with DST since 1999 where he leads a team working on object detection, including signal and information processing and data fusion.