Multi-Feature Based Multiple Pipelines Detection Using Ground Penetration Radar

K. Rajiv

Department of Computer Science and Technology, MRIET, Hyderabad, India.

Abstract

Ground penetrating radar (GPR) is one of the common sensor system used for underground inspection which emits electromagnetic waves that can pass through objects. The reflecting waves are recorded and digitized, and then, the images are formed. According to the properties of scanning object, GPR creates higher or lower intensity values on the object regions. Thus, these changes in signal represent the properties of scanning object. This technique includes transmission and reception of electromagnetic waves by means of which it is aimed at achieving an exploration depth of few meters with a resolution of several centimeters. For this interpretation of GPR data requires a human operation with high skill and experience, thus costing high amount of time and money. Therefore the development of noninvasive techniques to explore and retrieve information about the underground has shown a growing interest in the last years by public as well as private entities related to different application fields, such as oil and gas exploration, geology, conduits and pipes location. In this paper, an algorithm is composed of four steps; estimation of the number of multiple objects buried in the ground, isolation of each object, feature extraction and detection of pipelines. The number of objects in the GPR signal is estimated by using the energy projection method. Then signals for the objects are extracted by using the symmetry filtering method. Each signal is then processed for features, which are given as input for detection. Two pipelines buried in various ground conditions are considered for the test of the proposed method. They demonstrate that the proposed method can successfully detect buried pipelines.
I. INTRODUCTION

Ground-penetrating radar (GPR) is a geophysical method that uses radar pulses to image the subsurface. This non-destructive method uses electromagnetic radiation in the microwave band (UHF/VHF frequencies) of the radio spectrum, and detects the reflected signals from subsurface structures. GPR can be used in a variety of media, including rock, soil, ice, fresh water, pavements and structures. It can detect objects, changes in material, and voids and cracks. [1] Many efforts have been made for developing methods for detecting buried objects. Among them, detection using ground penetrating radar (GPR) has attracted many researchers’ attention due to its various advantages over other devices; it can be used to detect metallic and non-metallic buried pipelines [2] without changing its configuration. Moreover, it can be made into small and portable device so that it is used in various forms such as a stand-alone handheld sensor [3], a complementary sensor and vehicle-mounted system in the form of an array of multiple antennas. Detection of a buried pipeline from a GPR signal requires extracting features of a pipeline from the signal. There are various ways for computing features from the GPR signal for pipeline detection such as geometrical feature of a pipeline signal hidden Markov models (HMMs), Spatial Features analysis polynomial fitting, texture-feature coding method (TFCM), time-frequency features, and principal component analysis (PCA). The multiple features for a pipeline such as PCA and Fourier coefficients are used for robust detection and identification. Those methods are designed to detect a single pipeline from the GPR signal with some clutters. However, they are not intended to handle the case that multiple pipelines captured in one signal set. More than one pipeline may be buried in the ground, yielding a GPR signal containing multiple pipelines. In this situation, the methods for single pipeline detection are not able to operate properly if the signal is provided as input without a priori knowledge on the number of pipeline in the signal. It might happen that only one pipeline might be detected, and the others would be ignored as obstacles when the number of pipelines in the signal is not known. Although the implication of this problem is profound, little research on this problem has been presented in the literature.

In this work, the problem of multiple pipeline detection using the GPR is addressed, and a novel method is proposed, which detects multiple pipelines from a GPR signal. The GPR scans a ground, to generate a signal. Next, the number of objects in the signal that are determined to be possible pipelines is estimated. Once the number is computed, regions corresponding to each object are extracted and processed to obtain features. The features will help us to decide if the objects are pipelines or not.

II. METHODOLOGY

The overall procedure of multiple pipeline detection is proposed as shown in Fig. 1. It consists of two stages: segmentation and identification. The segmentation, which is indicated in a dotted rectangle in Fig. 1, estimates the number of possible pipelines in GPR data. This step is important in the multiple pipeline detection because it defines how many times the identification should be performed. Once the number of possible
pipelines is estimated, each segmented signal is processed for extracting features for identification. The identification, which is indicated in a black rectangle in Fig. 1, contains processes of signal extraction, signal post-processing and feature extraction. Once the features of each signal are computed, identification of the signal is performed. The first step is to normalize a GPR signal so that it orders to minimize the influence that might be caused by the difference of the hardware and the individual experimental environments.

Next, using the energy projection method the number of objects in the signal is estimated. Once the number has been estimated, by using the Symmetry Filtering Method signals for each individual object are extracted separately. To reduce noise and ground effects the extracted object signals are processed using the envelope detector and Gaussian filter. Then the features such as the principal components and geometric features are computed for each signal, and then provided as input to the identification module. The identification of the input signal is made using the support vector machine (SVM) based on the features.

III. MULTIPLE PIPELINE DETECTION ALGORITHM

3.1 Pre-processing of GPR Signal

A GPR signal is given as a set of intensity values at x and y positions as shown in Fig. 2. Here, the x axis indicates the integer index of each sampling position (denoted as Column No.) for the width of the scanning area. The sampling interval of the two adjacent scanning positions is 15 mm. The y axis is the depth and the z axis is the strength of the signal. The strength of the GPR signal reflects the size and material type of a pipeline in the ground. Therefore, it could be considered as a feature of a pipeline. However, the strength can also be influenced by other factors such as the
power of the GPR, the installation height of the radar and the properties of the ground. Therefore, the power intensities of B-scan signal are normalized for the strength in order to minimize such effects. The normalization is performed in a linear manner so that the strength of each signal is called to the range from zero to one.

![Fig.2. GPR signal in 3D space.](image)

### 3.2 Estimation of the Number of Objects

Estimation of the number of objects in the GPR signal mainly consists of three steps: removal of clutters, noise reduction and estimation of the number of objects. Assume that I is a GPR signal given as an image as shown in Fig. 3, where the strength of the signal is given as intensity values in the x-y (Column No. and Depth) plane.

![Fig.3. Example of an input signal given as an image in gray scale. The clutter due to the ground is clearly indicated.](image)

Clutters in the signal are removed from I using the average subtraction method [20], to yield Ic. The method is effective in removing the reflection from the ground, which is nearly constant in strength. It works as follows. I(a,b) is an intensity of the signal at
Multi-Feature Based Multiple Pipelines Detection Using Ground Penetration Radar

x and y. Then, the new $I_c(a,b)$, which is obtained by using the average subtraction method, is

$$I_c(a,b) = I(a, b) - E[I(a, b)]$$

Here, $E_a[]$ indicates the operator computing the average for all a at yi. This equation is applied at each yi. Fig. 4(a) shows an example of the signal $I_c$. It clearly shows that most of the ground reflection has been removed. Then, a $3 \times 3$ Gaussian mean filter with the sigma of 0.5.

$$f(a, b) = \left(\frac{1}{2\pi \sigma^2}\right)e^{\frac{-(a^2 + b^2)}{2\sigma^2}}$$

is applied to the signal $I_c$ for noise reduction, yielding the signal $I_{cs}$ of Fig. 4. Next, the Sobel edge detector [23] is applied to the signal $I_{cs}$ in order to extract signatures.

The detector is an algorithm to extract edges in an image using the gradient of the intensity distribution. Given the input intensity $I$=I(a,b).

$$\nabla I_m = \text{mag}(\nabla I) = \sqrt{\left(\frac{\partial I}{\partial a}\right)^2 + \left(\frac{\partial I}{\partial b}\right)^2}$$

If it is larger than or equal to a threshold, the intensity is set to one. Otherwise, it is set to zero. The threshold needs to be determined through various experiments. An example of the extracted edges is given in Fig. 4(c). The edges are then projected onto the y-z and x-z planes, respectively. During the projection, the intensity values at the same projected positions are accumulated to produce the accumulated projected signatures in the y-z and x-z planes. As demonstrated in Fig. 4(d), the projected values form distinct groups.

Then, the numbers of groups whose peaks are larger than a tolerance $T_2$ in each of the y-z and x-z planes are counted to be $n_{b2}$ and $n_{a2}$. Next, among $n_{b2}$ and $n_{a2}$, the numbers of groups, which have ranges (widths) wider than a tolerance $T_1$, are computed to be $n_{b1}$ and $n_{a1}$. The largest one between $n_{b1}$ and $n_{a1}$ is selected to be the number of objects in the signal, which is denoted by $n_{obj}$. This estimation procedure is called Energy projection method detected.
The results of the steps of the energy projection method. (a) The result of removing the constant clutter. (b) The result of applying the Gaussian filter. (c) The result of applying the Sobel edge detector. (d) The result of projecting GPR data to y-z (depth and intensity) plane.

T1 is the minimum width of the signal range in the projection planes. T2 is the minimum intensity in the GPR data. Signals either with narrower width than T1 or with intensity smaller than T2 are ignored as noise. Basically, these values need to be chosen empirically with various pipelines and objects. However, a guideline for the choice is that they should be so selected that the estimated number of objects in the signal becomes more than the actual number of pipelines. In this case objects other than pipelines can be discarded in the decision step. This way, the possibility that a pipeline is missed, a dangerous case than the false alarm, can be minimized.

3.3 Extracting Individual Objects

When an object buried in the ground is scanned by a GPR, a symmetric parabolic shape is obtained as shown in Fig. 4(a). Therefore, using a filter, which extracts a symmetric shape, a signal corresponding to the object in the ground can be obtained. The symmetry filtering method [21] can be used to handle this problem. The GPR data in this work are given as shown in Fig. 2. Then, the method can be applied as follows. First, the symmetry position in the signal is located using the following equation.

$$I_2(a, b) = \Sigma I(a - k, b - m)I(a + k, b - m)$$

Here, M and K are variables related with the radar pulse and the valid aperture of the radar given in . I is the intensity, and the values of x and y are the column number and
the depth, respectively. Next, the range direction symmetry weighting matrix is computed using the following equation.

\[ I_3 = D_m = (a - \mu)^T s^{-1} (b - \mu) \]

The function, \( \text{argmax}(I_2(x, y)) \), returns x coordinate, \( x_0 \), where \( I_2 \) in (4) is maximum. After that, the lateral direction symmetry weighting matrix is computed by

\[ I_4 = (I^a) + (I^b) \]

Once \( I_3 \) and \( I_4 \) are obtained, the synthetic symmetry filtering weighting matrix is computed by

\[ I_5(a, b) = e^{y s_n} e^{a, b} \quad (7) \]

Here, \( \gamma \) and \( \mu \) are the range and lateral reduction factors and set to one in this work. \( I_3n(y) \) and \( I_4n(x, y) \) are the normalized \( I_3(y) \) and \( I_4(x, y) \). The symmetric shape \( I_6 \) is then extracted by multiplying the input signal \( I \) by \( I_5 \).

\[ I_6(a, b) = I(a, b) I_5(a, b) \]

![Fig.5. The results of the symmetry filtering method.](image) (a) shows input data containing two symmetric patterns, \( P_1 \) and \( P_2 \). (b) is the result showing one symmetric pattern, \( P_1 \). (c) is the result of the other symmetric pattern, \( P_2 \) after \( P_1 \) has been removed.
Consider a GPR signal as shown in Fig. 5(a), where the clutter has been reduced and the number of objects has been estimated by using the procedure in Section 3.2. Next, the symmetry filtering method is employed to extract one symmetric pattern. The method detects the strongest signal in the data and extracts the associated symmetric pattern. As an example, the left one in Fig. 5(a) is extracted, which is denoted as $P_1$. The pattern is then subtracted from Fig. 5(a), to produce a signal without $P_1$, which is shown in Fig. 5(c). If the number of objects is estimated to be $n_{obj}$, this process is repeated $n_{obj}-1$ times to yield $n_{obj}$ signal sets, containing the symmetric patterns of each object.

### 3.4 Post-processing

Each extracted signal is processed to further reduce noise for feature extraction. A GPR signal changes its sign with respect to zero, due to the property of an electromagnetic wave. Such alternation is eliminated for the subsequent process using the Envelope Detection Method [15], which restores the original signal from those modulated in a high frequency. The detector is given by

$$y'_n = f_1x'_n + f_2y'_{n-1}e^{-\tau}. \tag{9}$$

Here,

$$f_1 = \begin{cases} 
1 & \text{if } x'_n \geq y'_{n-1}e^{-\tau} \\
0 & \text{otherwise}
\end{cases} \tag{10}$$

$$f_2 = \begin{cases} 
1 & \text{if } x'_n < y'_{n-1}e^{-\tau} \\
0 & \text{otherwise}
\end{cases}$$

Where $x(n)$ is the $n$-th input signal, $y(n)$ is the extracted envelope at position $n$ and $\tau$ is the reduction rate. The reduction rate needs to be selected such that the envelope detector yields the best result.

A large $\tau$ could reduce the effect of the envelope detector. On the other hand, a small $\tau$ may distort the original signal. In order to avoid such problems, a GPR frequency is considered such that the power of the signal reduces to half. Namely, an equation $e^{-\tau x} = 0.5$ is derived. From this equation, $\tau$ can be selected so that $x$ becomes equal to the GPR frequency. $y(0)$ is defined to be zero because a signal containing an object will not start at the beginning of the GPR signal. In other words, the distance between the GPR and an object is larger than zero.
Next, a Gaussian filter is applied in order to reduce high frequency noise in the signal. For this work, a Gaussian filter with a 3 by 3 window and the sigma of 0.5 is used. The result of the filtered signal is shown in Fig. 6. Next, for better feature extraction, the filtered signal is processed such that the intensity values, which are less than 50% of the maximum intensity in the signal, are set to zero. The result of this step is given in Fig. 7.

3.5 Extracting Features of an Object

The post-processed signal is used to extract features of a pipeline for identification. First, normalization of the signal is performed to consider the difference of the reflected power along the burial depth. The maximum and the minimum intensity values of the signal are scaled linearly to the range of 0 to 1, to adjust the power difference due to the different depths. Two geometric features and one statistical feature are considered. The two geometric features are the geometric dimensions of the signal intensity distribution as shown in Fig. 7. In the x-y (Column No. and Depth)
plane, intensity values larger than 50% of the maximum of the intensity usually leave an elliptic shape as shown in Fig. 7, which is obtained in the post-processing step given in Section 3.4. The shape can be characterized with the horizontal size and the vertical length. As a first feature the size, \( l_1 \) is measured along the horizontal axis as shown in Fig. 7. It is related with the size of an object reflected in the signal. The bigger an object is, the larger \( l_1 \) becomes. This feature can differentiate objects with respect to their size. This value is measured as the number of columns that the horizontal size of the region covers. The second feature is the length, \( l_2 \), which is measured along the depth (y axis) in the figure. It is mainly dependent on permittivity and permeability of materials of an object. The radio wave transmittance is different according to the object materials. It is observed that the length \( l_2 \) of a plastic object is longer than that of a metallic one. Therefore, this feature can separate a steel object from a plastic one. This feature is measured in mm.

As a third feature, a principal component by the principal component analysis (PCA) method is used. PCA method analyzes discrete data points in n-dimensional space, producing n principal components in n principal directions, forming n pairs of the component value and its corresponding direction. Each component represents a pattern of the data in that direction. The PCA has been frequently used in the context of detection, and its applicability can be extended to clutter reduction.

Consider \( m \) sets of data, \( r_i = (x_{1i}, x_{2i}, \ldots, x_{ni}), i = 1, \ldots, \min n \)-dimensional space. The average data of the \( m \) data sets are obtained by

\[
\Phi = \frac{1}{m} \sum_{i=1}^{m} \Gamma_i .
\]

Then the covariance of the data sets is then

\[
C_{\Phi} = \frac{1}{(m - 1)} \sum_{i=1}^{m} \varphi_i \varphi_i^T .
\]

The eigenvectors and eigen values of the covariance matrix are the principal components, which are computed by solving the following equation.

\[
C_{\Phi} \mathbf{u} = \lambda \mathbf{u} .
\]

Here, \( \mathbf{u} \) and \( \lambda \) are a vector and a scalar value, respectively. \( \mathbf{u} \) and \( \lambda \) satisfying (13) are the eigenvectors and their eigen values of the covariance matrix, serving as the principal components of the data.

A GPR signal is given in 3D space as shown in Fig. 2. The intensity distribution can be obtained by projecting the signal in the x-z and y-z planes. Among them, the intensity distribution of the signal projected onto the x-z (column no.vs. intensity) plane is selected because it shows clearer features of each pipeline. An example of
this projection is given in Fig. 8. Then, PCA is applied to the projected signal to obtain two principal components and their corresponding direction vectors as depicted in Fig. 8.

In this work, the direction close to the x axis happens to correspond to the largest covariance value because most of the signal values are clustered near the x axis in the x-z plane when the signal is projected on the plane. The eigenvectors are not suitable for features of an object because the signal patterns are almost symmetric so that they do not show any difference. The eigen value of Eigenvector 1 is not appropriate as a feature because most of the signal strengths are clustered near the x axis, making the eigenvector directed in the same axis. The eigen value of Eigenvector 2, however, captures the reflection patterns of an object, which depend on materials and the size of the object.

3.6 Pipeline Identification

Pipeline identification is performed by comparing features computed in Section 3.5 with those in a database that contains features of various pipelines buried in different ground conditions. If a match is found in the database, then the input pipeline is identified. There are a lot of algorithms for identification such as the hidden Markov models [6], Mahalanobis distance based method [14], the decision tree [12], Bayesian learning and Supporting Vector Machine (SVM) [12]. Among them, SVM is chosen in this work. It is a machine-learning algorithm used for data identification. A database is constructed, which contains features extracted from existing data sets. The features are classified into several groups, the outer most boundary elements of which are used to create optimal hyper planes differentiating each group. Algorithms for computing such hyper planes and robust decision are the main components of SVM. Once such a database is available, decision can be made that a group is determined where the features of input data belong. Various SVMs’ are introduced in the related
literature. In this work, a method using the radial basic function (RBF) is employed. The basic concept of SVM is presented as follows. Suppose that there exist various data groups, which can be distinguished from each other. SVM is an algorithm to determine the optimal boundaries of the data groups for the group classification. Consider there are two groups of data, A and B. Here, the circles are the data in Group A, and the triangles are the data in Group B.

![Fig.9. An example of SVM. (a) Shows two data groups with various separating boundaries. (b) Shows an optimal hyper plane separating the two groups. Here, \( l_m \) is the maximum margin.](image)

As shown in Fig. 9(a), an infinite number of boundaries can be introduced to separate the two groups. SVM selects one boundary that can optimally represent the boundaries of each group as illustrated in Fig. 9(b). In Fig. 9(b), the dotted line is the optimal separating line, called the optimal hyper plane, which is determined in such a way that the distance between the two data groups becomes maximal, or the maximum margin, \( l_m \), becomes maximal. In general, SVM can be formulated as follows. Consider that there exist data \( D_s \) in \( p \) dimensions.

\[
D_s = \{(x_i, y_i) | x_i \in \mathbb{R}^p, y_i \in \{-1, 1\}, i = 1, \cdots, n_p\}.
\]  

(14)

Here, \( x_i \) is a vector in \( \mathbb{R}^p \), \( n_p \) is the number of data, and \( y_i \) isthe vector class. The hyper plane of this data set is obtained by solving the optimization problem:

\[
\min \frac{1}{2} w^T x + c \sum_{i=1}^n \xi_i
\]

subject to:

\[
y_i(w^T x_i + b) \geq 1 - \xi_i, j = 1, \cdots, l.
\]

(15)

Here, \( w^T \) is the normal vector of a hyper plane, \( x \) is the input data set, \( \xi \) is a slack variable and \( C \) is a constant for providing a constraint to Lagrange multipliers.

Given the sign function,
a decision function using the hyper plane for determination of a data class is given as:

$$f'(x) = \text{sgn}(wx + b).$$  \hfill (17)$$

In this work, three features are considered as discussed in Section 3.5. Each feature corresponds to one axis, forming 3D feature space. Therefore, one object is represented as one point in the feature space as shown in Fig. 10. As illustrated in the figure, a pipeline buried in the ground of different conditions and at various depths forms a cluster in the space. These data are provided as input to SVM for constructing a database. Later, given a point in the space, detection of the point is performed using the database.

![Fig.10. A plot of features of two pipelines buried in dry sand.](image)

In order to determine if an object is a pipeline or not, the three feature values for objects other than pipelines, so called clutters, are added to the data base to form groups for non-pipelines. Various rocks different in size or plastic and metal cans can be processed for this purpose. The database needs to be updated to contain features of various objects, which would increase the identification rate for pipelines.

### 3.7 Implementation

The proposed method was implemented using a PC of 2.7 GHz CPU and 8 Gbyte RAM. The operating system of the computer was Windows 7, and Visual C++ and
Matlab were used for implementation. Conceptually the proposed method consists of six primary modules, modules for pre-processing, estimating the number of objects, extracting objects’ signals, post-processing, extracting features and identification. Each primary module consists of one or more functions. The preprocessing module contains a function for normalization. The module for estimation of the number of objects contains a function of the energy projection method. The module for signal extraction provides a function of the symmetry filtering method. The post-processing module is composed of two functions: envelope detection and Gaussian filter. The feature extraction module consists of three functions of computing PCA and geometric lengths. The theoretical basis of these functions is given in Section 3.5. The identification module contains a function of SVM for identification as given in Section 3.6. a core part of the implementation, based on which the functions for processing GPR signals and identification are implemented. The functions have interfaces with the file I/O and data handling for input and output. The sequence of execution of the functions is managed by the file I/O and data handling function.

IV. CONCLUSIONS

This paper proposes a novel method for detecting multiple pipelines buried in the ground of various conditions using a GPR. The method consists of clutter reduction, estimation of the number of objects in the GPR data, isolation of object signals, feature extraction and detection. The tests show that the proposed method can mostly detect multiple pipelines in various conditions. The proposed method has two limitations: choice of tolerances and database construction. The method requires a couple of user-defined values: threshold values for the estimation of the number of objects and a parametric value used by the symmetry filtering method. These values mainly depend on the properties of the GPR hardware used in the system and are critical in the process. The other limitation is that a database for pipelines and various other foreign objects should be constructed for robust detection. The database for non-pipelines could be limited in that it cannot consider all possible clutters different in size, shape and material. It means that there might be a false-alarm case that a non-pipeline object is determined as a pipeline due to the limited amount of data in the database. These two limitations can be overcome through extensive experiments with objects and pipelines. The overall procedure of the method is designed for general environments. However, the tests in the paper were performed in controlled conditions, and the tolerances were chosen and the database was constructed for those conditions. Therefore, it is necessary to refine the tolerances and to improve the database by considering more realistic situations. A thorough evaluation of the method for selection of tolerances and enhancing the database using real field data is recommended for future work.
REFERENCES


BIBLIOGRAPHY

Mr. K. Rajiv did his M.Tech in Computer Science and Engineering from SRKR Engg. College, Bhimavaram. Presently, he is working as Assistant Professor in the Department of CSE, MRIET, Hyderabad. He is currently pursuing his PhD in Acharya Nagarjuna University, Guntur. He published 5 papers in National and International Journals and 3 papers in International Conferences. He has 6 years of Teaching Experience.