Mapping the Buried Cable by Ground Penetrating Radar and Gaussian-Process Regression

Xiren Zhou, Qiuju Chen, Shengfei Lyu, Huanhuan Chen, Senior Member, IEEE

Abstract—With the rapid expansion of urban areas and the increasing use of electricity, the need for locating buried cables is becoming urgent. In this paper, a novel method to locate underground cables based on Ground Penetrating Radar (GPR) and Gaussian-process regression is proposed. Firstly, the coordinate system of the detected area is established, and the input and output of locating buried cables are determined. The GPR is moved along the established parallel detection lines, and the hyperbolic signatures generated by buried cables are identified and fitted, thus the positions and depths of some points on the cable could be derived. On the basis of the established coordinate system and the derived points on the cable, the clustering method and cable fitting algorithm based on Gaussian-process regression are proposed to find the most likely locations of the underground cables. Furthermore, the confidence intervals of the cables’ locations are also obtained. Both the position and depth noises are taken into account in our method, ensuring the robustness in different environments and equipment. Experiments on real-world datasets are conducted, and the obtained results demonstrate the effectiveness of the proposed method.

Index Terms—Ground Penetrating Radar (GPR), buried asset detection, pipeline mapping, Gaussian process regression.

I. INTRODUCTION

Underground cable plays an important role in the energy transmission of urban facilities, including probes, street lights, and power exchange stations. Locating underground cables precisely has always been a prerequisite for maintaining the normal operation of the urban power system [1]. Existing pipeline maps could generally provide rough locations of the buried cables, but it’s challenged to be specific to each point on these cables [2]. Therefore, it is vital to locate buried cables before excavation and construction. Unlike unbendable underground pipes whose location could be described by straight line segments, locating a buried cable could be abstracted into locating an underground curved segment [3], which could be divided into two parts: determine the position and depth of some points on the buried cable, and then estimate the location of the whole cable based on these points.

To extract information of buried facilities in the shallow subsurface, Ground Penetrating Radar (GPR) has been widely used due to its non-destructive property [4]. If the cable or pipeline is buried within the effective detection depth of the GPR and has a different dielectric constant from the surrounding medium, a hyperbolic signature would be formed on the obtained GPR B-scan image after moving the GPR across the cable or pipeline [5]. By identifying and fitting the hyperbolic signature on the B-scan image, the position and depth of the buried cable or pipeline on the cross-section GPR moved could be estimated [6].

There are published methods to identify and interpret hyperbolic signatures on B-scan images, including Hough-transform-based methods [7], [8], least-square methods [9]–[12], machine-learning-based methods [13]–[19]. Some combinations of the above kinds of methods are also conducted to obtain more precise results [20]–[24]. In [21], Chen et al. proposed a probabilistic hyperbola mixture model to extract and fit hyperbolic point sets on the B-scan image. The Expectation-Maximization (EM) algorithm is upgraded to extract points on multiple hyperbolic signatures from a GPR image, which is then fitted to estimate the equation of each hyperbolic point set. In [23], the position and time data are paired and recorded in a generalized pair-labeled Hough transform, and a conventional least-square method is then utilized to infer the position, depth, and radius of the buried pipeline. In [24], the Column-Connection Clustering (C3) algorithm is proposed to scan the B-scan image and separate point sets on intersections. Point sets with hyperbolic signatures are then identified by a neural-network-based method. In our previous work [25], a GPR B-scan image interpreting model has been proposed. The model could estimate the radius and depth of the buried pipelines by converting the GPR B-scan images into binary ones, scanning the binary image to cluster points, and fitting the obtained point sets with hyperbolic signatures. Experiments on cement and metal pipes have been conducted and the results validated the accuracy and efficiency of this model. Subsequently in [26], the model is further extended to estimate insulated pipes in the soil by combining GPR with Electric-Field (EF) methods. By applying the above methods, the depth and position of some points on the buried cable or pipeline could be roughly derived from GPR B-scan images.

Besides GPR, the Electromagnetic Induction (EMI) is also widely used technology for estimating buried facilities [27]. EMI consists of two steps. A transmitter is firstly used to transfer an alternating electrical current to the pipe or cable to be located either by a direct connection [28] or by inducing a signal [29]. Next, a receiver is used to receive the transmitted signal. The received signal could then be analyzed to localize the position of the buried facility [30]. The use of the EMI method has requirements on the metallic material and continuity of the target facility. In real-world applications,
both EMI and GPR’s detection accuracy could suffer from the aboveground and underground noises [4], [31]. Thus the obtained data should be further denoised and processed to obtain more accurate and reliable results.

Once the cable information at some points is derived, the following work is to infer the locations of the cables in the detected area from these individual hypothesized points. Except for the noise and errors that occur in the interpretation of GPR data, the positioning errors could not be ignored either. In the case of tall buildings and trees, satellite-based positioning signals could be blocked [32]. The positioning accuracy of an odometer would degrade when a rough or slippery ground is measured [33], and the errors would accumulate using this kind of methods as the detection distance increases. In [34], the directions of buried objects are roughly determined by the existing pipeline map. Multiple GPR detections in different directions are then conducted, and the obtained hyperbolas are connected to derive the specific location of each buried object. In [3], a Three-Dimensional Spline Interpolation (TDSI) algorithm is proposed, where detected points on the cable obtained from GPR are interpolated with a smooth curve. The main concern of these two algorithms is that both positioning and detecting errors are not taken into account to obtain more reliable location of buried objects. In [35], a Marching-Cross-Sections (MCS) algorithm is proposed to merge the individual hypothesized pipeline information from multiple sensors and locate underground pipeline segments. Detected points on the pipelines or cables are connected by an extended Kalman Filter (KF) [36] with straight line segments, assisted by some rules that manage buried utilities to keep potentially existing ones and to discard invalid ones. In [37], a probabilistic mixture model is conducted to denoise and classify the data from detected points, which are then fitted by a Classification Fitting Expectation Maximization (CFEM) algorithm to locate the buried pipeline. The above methods [35]–[37] are mainly aimed at mapping buried pipelines that are straight. However, in order to avoid underground obstacles, buried cables might be bent, as well as the use of pipe-jacking technology1 [38]. Thus describing the locations of buried cables by straight line segments might lead to inaccurate results.

In addition to the above methods, there are also some published work that estimates the orientation of a subsurface linear target through polarimetric GPR data. In [39] and [40], the linearity factor is used to classify rotationally symmetric and linear objects, which is estimated from the eigenvalues of the scattering matrix obtained from polarimetric GPR data. A linear target has a preferential scattering axis that coincides with its long axis [41]. The strike direction of a subsurface fracture could be estimated by minimizing the energy of the cross-polarized components, which is realized by zeroing the derivative of the energy in the cross-polarization channel [42]. In [43], the orientation of the buried linear target is estimated by maximizing the energy in the two cross-polarization channels, and the result is then obtained from the average of two angles. In [44], a hybrid dual-polarimetric GPR system, which consists of a circularly polarized transmitting antenna and two linearly polarized receiving antennas, is employed to estimate buried linear objects. A full-polarimetric scattering matrix is extracted from the double-channel GPR signals reflected from a buried linear object, and then an improved Alford rotation method is utilized to estimate the orientation angle of the object from the extracted scattering matrix. The aforementioned methods identify the orientation of linear targets mainly from the scattering matrix extracted from the polarimetric GPR data. For given equipment and data, such as the GSSI’s SIR-30 GPR utilized in this paper and the obtained B-scan images, effective methods are still needed to accurately locate buried cables within the detected area.

To address the above issue of localizing buried cables in the presence of noises, a novel method to locate buried cables is proposed in this paper. The input/output of locating the buried cables is determined by conducting a coordinate system with parallel detection lines2, along which the GPR is moved to collect B-scan images. The established detection lines are independent of each other, and the positioning error on one detection line will not affect other detection lines, reducing the error accumulation with detecting distance. The obtained images are denoised and processed to highlight and extract hyperbolic point sets by extending part of our previous work [25], which are fitted to estimate the positions and depths of some points on the buried cables at each detection line. Instead of connecting the obtained points with straight line segments or interpolating them with curves, these points are clustered and processed by a designed cable fitting algorithm based on Gaussian-process regression to obtain the most likely locations of the underground cables. In the proposed method, both the position and depth errors are taken into account. The cable’s locations obtained at detected points are correlated and modified to infer the cable’s location within the detected area. The confidence interval of the buried cable is also obtained to infer the range where the cable is most likely to exist, which could provide early warning for the excavation, and could also reduce the range for further precise detections. The main contributions of this paper could be summarized as follows:

1) The input/output of locating buried cables is determined by conducting a coordinate system of the detected area, normalizing the data set composed of detected points, and describing the locations of the cables. Independent parallel detection lines are established based on the conducted coordinate system, where the positioning error on one detection line will not affect detections on other detection lines.

2) A cable fitting algorithm based on Gaussian-process regression is proposed, which takes both depth and position errors into account, improving the robustness of the proposed method. The algorithm could provide the most likely locations of the buried cable represented as a curve segment, and the confidence intervals of the buried cables are also obtained. The measured cable positions in the conducted experiments are all within the obtained confidence intervals.

1Pipe-jacking is a trenchless technology for buried utilities.

2A detection line indicates a line segment on a ground surface where the GPR is moved to collect the B-scan image.
3) A hyperbolic fitting algorithm is designed for the signature on B-scan images generated by cables based on the denoising and hyperbolic point set extracting method proposed in our previous work [25]. On the basis of Restricted Algebraic-Distance-based Fitting algorithm (RADF) [25], the hyperbolic point sets are fitted with more accurate equations through several iterations that minimize the sum of orthogonal distances.

The procedure of the proposed method of locating buried cables is presented in Fig. 1.

The rest of this paper is organized as follows. The GPR B-scan image interpreting model is introduced in Section II. Section III describes the method of locating buried cables based on Gaussian-process regression, including conducting the coordinate system, clustering the detected points into different cables, and finding the most likely locations and confidence intervals of the buried cables. Experiments are conducted and analyzed in Section IV. Finally, conclusions are drawn in Section V.

II. THE GPR B-SCAN IMAGE INTERPRETING MODEL

In this section, the theoretical model of estimating buried cables by GPR is presented. Then the GPR B-scan image interpreting model that extends part of our previous work [25] is proposed. The fitting result of Restricted Algebraic-Distance-based Fitting algorithm (RADF) [25] is used as the initialization, and a more accurate hyperbolic equation is obtained by iteratively minimizing the sum of the orthogonal distances from points to the target hyperbola.

A. The Hyperbolic Model Generated by the Buried Cable

The hyperbolic signatures on the GPR B-scan image are often formulated as a geometric model, where the signal assumes the diagram of a function of the GPR position on the detected line to the two-way travel time of the electromagnetic magnetic wave [45].

\[ \left( \frac{v_f}{2} \right)^2 - (y_i - y_0)^2 = \left( \frac{v_f t_0}{2} \right)^2, \]

which could be converted to

\[ \frac{t_i^2}{t_0^2} - \left( \frac{y_i - y_0}{v_f} \right)^2 = 1. \]

This is a hyperbolic equation about \( t_i \) and \( y_i \). By fitting the point sets with hyperbolic equation, the position and depth of the buried cable could be derived [3].

B. Interpreting GPR B-scan Images

Interpreting the GPR B-scan image to estimate buried cable consists of two parts: extracting point sets with hyperbolic signatures, and fitting the extracted point sets to estimate the cables’ position and depth.

1) Extracting Point sets with Hyperbolic Signatures: When utilizing GPR for underground cable detection, some operations are conducted to eliminate the noise and highlight the targets, which consists of three tasks: 1) Eliminating the undesired presence of the ground surface echo; 2) Reducing background noises; 3) Compensating propagation losses. The reflectance of the ground surface is firstly eliminated in advance. This work is supported by the Matgpr\textsuperscript{4} [46]. After that, a filtering step based on the standard median filter [47] is performed to reduce the electromagnetic noise and interferences. Finally, concerning the compensation of the

\textsuperscript{3}The two-way travel time represents the time that the wave runs from the transmitter to the object then to the receiver.

\textsuperscript{4}Matgpr is a freeware matlab package for the analysis of common-offset GPR data.
propagation losses caused by the medium attenuation and the signal energy radial dispersion, a nonlinear time-varying gain is applied to the received signal [48].

Based on the above operations, the GPR B-scan image preprocessing method, Open-Scan Clustering Algorithm (OSCA), and Parabolic Fitting-based Judgment method (PFJ) are utilized to extract hyperbolic point sets generated by the buried cable. The preprocessing method transforms GPR B-scan images into binary images, and removes most of the discrete noisy points. In the preprocessing method, the GPR B-scan image is firstly thresholded into the binary image based on the gray value on the boundary, and then the opening and closing operations [49] are applied to smooth the contour of the objects, eliminate small protrusions and fill small gaps in the contour noisy points. After preprocessing, the Open-Scan Clustering Algorithm (OSCA) is utilized to scan the binary image progressively from top to bottom, finds the openings, and conduct clustering. Point sets with downwardly-opening signatures are distinguished and extracted even if there are intersections and interference between them, and the point sets without downwardly-opening signatures are eliminated. After that, Parabolic Fitting based Judgment method (PFJ) is applied to further identify point sets with hyperbolic signatures. The preprocessing method, OSCA, and PFJ were proposed and introduced in detail in our previous work [25], thus they would not be detailed in this paper. An example of the process of extracting point sets with hyperbolic signatures is illustrated in Fig. 3. The obtained point set would be fitted by the hyperbolic fitting algorithm introduced in follows.

![Image](a) (b) (c) (d)

Fig. 3. These four images show the processing flow of extracting a hyperbolic point set from a GPR B-scan image. (a) is the B-scan image processed by operations to eliminate the noise and highlight the targets. (b) is the thresholded binary image. (c) is the binary image after opening and closing operations, where the contour of the objects is smoothed, small protrusions are eliminated, and small gaps in the contour noisy points are filled. (d) is the obtained result of OSCA and PFJ, where the point set with hyperbolic signature is identified and extracted.

2) Fitting the Extracted Point Sets: The fitting algorithm could be divided into two steps. Firstly, the RADF [25] is used to quickly fit the point sets to Equation (2) as the initialization. After that, the initial hyperbolic equation is modified to obtain a more accurate result by minimizing the sum of orthogonal distance from points to the target hyperbola through several iterations.

The general hyperbola with the focal point on the vertical axis could be presented as

\[
\frac{(z - z_0)^2}{A^2} - \frac{(y - y_0)^2}{B^2} = 1. \tag{3}
\]

Relating Equation (3) and (2), it could be seen that \( z_0 = 0 \) in the hyperbola generated by the cable. Thus the parametric form the hyperbola could be presented as

\[
\begin{align*}
y &= A \sinh \varphi + C, \\
z &= B \cosh \varphi,
\end{align*}
\tag{4}
\]

where \( C = y_0 \). Given a point set \((y_i, z_i)_{i=1}^m\), the orthogonal distance \( d_i \) from a point \( p_i = (y_i, z_i) \) to the hyperbola could be expressed by

\[
d_i^2 = \min \{ (y_i - y(\varphi_i))^2 + (z_i - z(\varphi_i))^2 \},
\tag{5}
\]

where the point \((y(\varphi_i), z(\varphi_i))\) is the nearest corresponding point of \( p_i \) on the hyperbola. The \( A, B \) and \( C \) could be determined by

\[
\min \sum_{i=1}^{m} d_i^2,
\tag{6}
\]

which is equivalent to solving the nonlinear least squares problem

\[
\begin{align*}
(y_i) - (C) - (A \sinh \varphi_i) \\
(z_i) - (B \cosh \varphi_i) \\
\approx 0, \text{ for } i = 1, \cdots, m.
\end{align*}
\tag{7}
\]

Thus, there are \( 2m \) nonlinear equations for \( m + 3 \) unknowns: \( \varphi_1, \varphi_2, \cdots, \varphi_m, A, B, C \). To solve the minimization problem, the Gauss-Newton iteration is employed. As aforementioned, the initialization of the Gauss-Newton iteration is determined by RADF. The experimental studies in Section IV demonstrate that this initialization is appropriate and robust against noise. Moreover, in the conducted experiments, the accuracy of the proposed fitting algorithm is also better than RADF when fitting hyperbolic signatures generated by buried cables.

III. LOCATING BURIED CABLES BASED ON GAUSSIAN-PROCESS REGRESSION

In this section, the coordinate system is established with parallel detection lines. The input and output of locating buried cables are also normalized. Then the clustering method and cable fitting algorithm are applied to estimate the location and confidence intervals of the buried cables in the detected area, provided the depths and positions of detected points at each detection line. As other probabilistic methods [50]–[52], one of the benefit to use Gaussian-process regression is the probabilistic outputs with confidence.

A. The Coordinate System and the In/Output of Locating Buried cables

The conducted coordinate system is visualized in Fig. 4, where the \( Y \) axis indicates the direction where GPR is moved (parallel to the detection line), and the direction of \( X \) is perpendicular to \( Y \). The downward direction perpendicular to the horizontal \( XOY \) plane is set to be the positive direction of \( Z \) axes. Multiple parallel detection lines are established (the red line in Fig. 4), where the GPR is moved to obtain B-scan images (a gray surface in Fig. 4 represents a GPR B-scan image).

The GPR B-scan image obtained at each detection line is processed by the model introduced in Section II. If a hyperbolic point set generated by a cable is identified and fitted, the top of the hyperbola is a detected point on the
cable [3], and the coordinates on Y and Z of this point are recorded. The coordinate on X axis is obtained from the distance between the detection line to the origin O. Based on the above, the cable’s location in a detection line could be expressed by a detected point \((x_i, y_i, z_i)\) that is made of the coordinates on the three axes, and the cable’s location obtained from all the detection lines would compose a point set as

\[
P = \{ P_i (x_i, y_i, z_i) | 0 \leq i \leq n \},
\]

where \(P_i\) is the number of detected points. It should be noted that when \(i \neq j\), \(x_i\) might be equal to \(x_j\), since there could be more than one detected cable in a GPR B-scan image obtained at a detection line. As the detection line gradually moves away from the origin \(O\), the coordinates on \(X\) axis of the acquired detected point on each detection line are increasing, which means for \(i < j\), there is \(x_i \leq x_j\).

Based on the conducted coordinate system, the location of a buried cable could be described by two respective functions that map any real number \(x \in [x_0, x_n]\) on the \(X\) axis to \(y\) and \(z\) on the \(Y\) and \(Z\) axes. The location \(L\) of the buried cables in the detected area could be described as:

\[
L = \{(f_{yc}(x), f_{zc}(x), \sigma_c(x)) | c \in [1, C], c \in \mathbb{N}, x \in [x_0, x_n], x \in \mathbb{R}\},
\]

where \(C\) is the number of cables in the detected area, \(f_{yc}(x)\) and \(f_{zc}(x)\) indicate the two functions of the \(c\)th cable, which map any real number \(x \in [x_0, x_n]\) on the \(X\) axis to corresponding \(y\) and \(z\) on the \(Y\) and \(Z\) axis, \(\sigma_c(x)\) is the confidence interval at any \(x \in [x_0, x_n]\), which indicates the minimum interval of each \(x\) where the pipeline existence probability is greater than or equal to 95%.

### B. Clustering Detected Point into Different Cables

The proposed GPR B-scan image interpreting model could estimate more than one cable in an GPR B-scan image, thus detected points from several cables could be obtained at a detection line. As the bending of the underground cables should be limited, otherwise the cable might be damaged [3], thus the detected point \((x_i, y_i, z_i)\) of the cable at \(l\)th detection line could be selected by:

\[
\min \left( \frac{\vec{D}_i \cdot \vec{D}_j}{|\vec{D}_i| \times |\vec{D}_j|} \right),
\]

where \(\vec{D}_i = (y_i - y_j, z_i - z_j)\), and \(y_i, z_i\) are from the points on the \((l - 1)\)th detection line. For the first detection line, the cable’s direction could be assumed to be perpendicular with the detection line, thus \((x-1, y-1, z-1) = (-1, y_i, z_0)\).

Equations (10) is applied to process the detected point from the detection line closest to the origin \(O\), line by line along the \(X\) axis. Each cable could generate only one detected point in a detection line. If a detected point is obtained in a detection line that does not belong to any previous cable, a new cable will be created and the direction of this cable is set to be perpendicular with the detection line. When all the detected points are processed, the cable with too short length are discarded, since there might be underground objects incorrectly identified as the buried cable (for example, when the distance between two nearest detection lines is about one meter, the cable segment that contains only one or two detected points should be discarded). After that, the partition of \(P\) is created as

\[
P_c = \{ P_c | c = 1, 2, \cdots, C \},
\]

in which \(P_c\) indicates the set of detected points generated from the \(c\)th cable.

### C. Cable Fitting Algorithm Based on Gaussian-Process Regression

Considering the existing noises of position and depth when detecting buried cables, using a pre-defined function [3] to fit the detected points on a cable could be inaccurate. Applying a function with the maximum likelihood is more tolerant in this situation. In this paper, the Gaussian-process regression is expanded to fit detected points on the cable. The cable’s locations obtained at detected points are correlated and modified in this process to infer the cable’s location within the detected area and reduce the impact of errors, where noises on \(Y\) and \(Z\) axes are assumed to follow an independent normal distribution, respectively. In this process, the error of a detected point could only affect the fitting result within a limited distance range without accumulation. The most likely location of the buried cable is obtained through this process. Also, the confidence interval of the cable’s location could be obtained to infer the range where the cable is most likely to exist.

As aforementioned, \(P_c\) indicates a set of detected points generated by a cable, which is a subset of \(P\). It is assumed that the horizontal location \(y_i\) and the vertical location \(z_i\) are independent from each other but both related to \(x_i\), thus \(P_c\) could be separated into two data subsets:

\[
P_{c1} = \{(x_i, y_i) | 0 \leq i \leq n\},
\]

\[
P_{c2} = \{(x_i, z_i) | 0 \leq i \leq n\}.
\]

These two two-dimensional subsets \(P_{c1}\) and \(P_{c2}\) are fitted separately, and then integrated into the location information of the entire cable in three-dimensional space.

A Gaussian process indexed by \(x = x_0, x_1, \cdots, x_n\) (time or space) is a stochastic process such that every finite collection of random variables has a multivariate normal distribution, and it could be specified by its mean function \(m(x)\) and covariance function \(k_f(x, x_j)\), \((0 \leq i, j \leq n)\). Intuitively, a real-valued
function \( f(\cdot) \) could be viewed as a vector with infinite dimensionality, and could be sampled from a normal distribution that specifies the function space. Therefore, Gaussian processes define the prior distribution of a latent function \( f(\cdot) \), and could encode the assumptions of \( f(\cdot) \) in the design of covariance function \( k_f(\cdot, \cdot) \) instead of choosing any specific form of \( f(\cdot) \).

For the point set \( P_{c1} = \{(x_i, y_i)|0 \leq i \leq n\} \), the Gaussian-process regression model could be described as:

\[
y_i = f(x_i) + \epsilon_i, \quad i = 0, 1, \ldots, n,
\]

where \( \epsilon_i \) indicates the noise variable with independent normal distribution, then a prior distribution over functions \( f(\cdot) \) could be assumed:

\[
f(x) \sim \mathcal{GP}(m(\cdot), K(x, x)),
\]

where \( x = (x_i)_{i=1}^n \), \( m(\cdot) \) is the mean function, and \( K(\cdot, \cdot) \) is covariance matrix satisfying \([K(x, x)]_{ij} = k_f(x_i, x_j)\). The mean function \( m(\cdot) \) could be assumed as zero to reduce the amount of parameters [53]. The covariance function \( k_f(x_i, x_j) \) maps the distance between \( x_i \) and \( x_j \) to the covariance between \( f(x_i) \) and \( f(x_j) \), which forms as:

\[
k_f(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2}(x_i - x_j)^2\right) + \theta^2 \delta_{ij},
\]

where \( \theta \) is a hyperparameter, and \( \delta_{ij} \) is a Kronecker delta which is one if \( i = j \) and zero otherwise. The covariance function indicates that the closer two detected points are, the higher correlation they have. This is justified since the actual location of target point on the cable is more related to the nearby points, and Equation (16) reduces the noises from the detected points in the distance. In the case of the same distance, the smaller the \( \theta \), the larger the covariance, which means that the cable position at this point is more correlated with the surrounding cable positions. In actual engineering, \( \theta \) is affected by the material of the cable. The greater the bendability of the buried cable material, the smaller the correlation between a point on the cable and the surrounding points, since this cable could be arbitrarily bent or change direction. Considering the existence of noises in detecting, the term \( \theta^2 \delta_{ij} \) is added to the function to follow the independence assumption about the noise from the positioning method or the depth information obtained by interpreting B-scan images.

To calculate the location \( y^* \) for a new position \( x^* \), it is assumed that \( y^* \) and all regression targets \( y \) in \( P_{c1} \) has the same joint normal distribution with zero mean function, where \( y = (y_i)_{i=1}^n \). Then we have:

\[
\begin{pmatrix} y \\ y^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K & k_{c^*T} \\ k_c^* & k_{c^*c} \end{pmatrix}\right),
\]

where \( k_{c^*} \) is a covariance vector in which the \( i \)th element is \([k_{c^*}]_i = k_f(x^*, x_i)\) and \( k_{c^*c} = 1 \), and \( k_{c^*T} \) is the transposed matrix of \( k_{c^*} \). Then the distribution of \( y^* \) could be calculated via Equation (17) as:

\[
y^* \sim \mathcal{N}(\mu^*, \sigma^*).
\]

The mean value \( \mu^* \) and variance value \( \sigma^* \) of target \( y^* \) could be calculated through Equations (19) and (20) as:

\[
\mu^* = k_{c^*T} K^{-1} y, \quad (19)
\]

\[
\sigma^* = -k_{c^*T} K^{-1} k_{c^*} + k_{c^*c}, \quad (20)
\]

where \( \mu^* \) indicates the mean value of the output of all valid functions in which these functions fit the detected data points, and \( \mu^* \) serves as the most likely location of the buried cable. Considering the existence of noises in the detected points, \( \sigma^* \) presents the confidence interval by \( \mu^* \pm 2\sigma^* \).

The above process is applied to both \( P_{c1} \) and \( P_{c2} \) separately to obtain the location on \( Y \) and \( Z \) axis, and then combined to describe the three-dimensional location of the buried cable. For each cable, given an \( x \) in the detected area, the corresponding \( y \) and \( z \) could be obtained. When mapping the buried cable, the noise of \( y \) mainly comes from the positioning error, and the main source of the noise of \( z \) is the error produced from interpreting the GPR B-scan image to obtain the cable’s depth. The noises of \( y \) and \( z \) in our method are regarded as independent, and the hyperparameter \( \theta \) in the processing of \( P_{c1} \) and \( P_{c2} \) are decided separately as \( \theta_y \) and \( \theta_z \), which depends on the intensity of the two kinds of noises.

### D. The Pseudo Code of Locating Buried Cables

The pseudo code of clustering and fitting the obtained detected point set is presented as Algorithm 1.

#### Algorithm 1 Locating buried cables by Gaussian-process regression

**Input:** Detected points set \( P = \{(x_i, y_i, z_i)|0 \leq i \leq n\}, \beta \) and \( \theta \).  

**Output:** The most likely locations and confidence intervals of all buried cables \( L = \{(f_{yc}(x), f_{zc}(x), \sigma_c(x))\} \).

1: Cluster the points in \( P \) into independent subset \( P_c (c = 1, 2, \ldots, C) \), where \( P_c \) consists of detected points generated by the \( c \)th cable, and \( n_c \) indicates the number of detected points in \( P_c \).

2: for every \( P_c \) in \( P \) do

3: Separate \( P_c \) into \( P_{c1} = \{(x_i, y_i)|0 \leq i \leq n_{c1}\} \) and \( P_{c2} = \{(x_i, z_i)|0 \leq i \leq n_{c2}\} \).

4: Calculate the covariance matrix \( K_c \) on \( x_c \), where

\[
k_f(c, c) = \exp\left(-\frac{1}{2\sigma^2}((c - c)^2) + \theta^2 \delta_{cc}\right).
\]

5: for every \( x_{c1} \in [x_{c1,0}, x_{c1,n}] \) do

6: for every \( x_{c2} \in [x_{c2,0}, x_{c2,n}] \) do

7: Calculate the covariance between \( x_{c1} \) and \( x_{c2} \):

\[
k_{c_{c1}}(x_{c1}, x_{c2}) = \exp\left(-\frac{1}{2\sigma^2}((x_{c1} - x_{c2})^2)\right).
\]

8: end for

9: \( k_c = \{k_{c, c} | x_c \in x_c\}, \)

10: The function of the most likely locations \( f_{yc} \) and \( f_{zc} \):

\[
y_c = (y_c|0 \leq i \leq n_{c1}, z_c = (z_c|0 \leq i \leq n_{c2})
\]

\[
f_{yc}(x_{c1}) = k_{c, c}^* K_c^{-1} y_c,
\]

\[
f_{zc}(x_{c2}) = k_{c, c}^* K_c^{-1} z_c.
\]

11: The confidence interval:

\[
\sigma_c(x_{c1}) = -k_{c, c}^* K_c^{-1} k_{c, c} + 1,
\]

\[
f_{yc}(x_{c1}) \pm \sigma_c(x_{c1}), \quad f_{zc}(x_{c2}) \pm \sigma_c(x_{c1}).
\]

12: end for

13: end for

14: return \( L = \{(f_{yc}(x), f_{zc}(x), \sigma_c(x))|c \in [1, C], c \in N, x \in [x_{c1,0}, x_{c1,n}]\} \).
IV. EXPERIMENTAL STUDY

In this section, experiments on real-world datasets are conducted. After that, the analysis of the experimental results and some comparative work are presented.

A. Experimental Environment and Settings

By consulting the existing underground pipeline map, three experimental areas are identified. The existing piping map in these areas provide some points on the cables, which are connected by straight lines. GSSI’s SIR-30 GPR with 200-MHz antenna is utilized to collect GPR B-scan images, and the GPR’s supporting positioning equipment would record the position of every point along the detected path. When a hyperbola is identified and fitted, the position of the buried cable at this point could be obtained and recorded. The three selected areas and the utilized GPR are visualized in Fig. 5. These areas are all near the roads with electric utilities nearby, such as cameras, street lights, etc. Part of the buried cables in these areas are evacuated as Fig. 6, which demonstrates the fact that underground cables could not be accurately located by straight line segments.

![Figure 5](image1.png)

![Figure 6](image2.png)

Fig. 5. The satellite maps of three selected areas and the established coordinate systems are shown as (a), (b), and (c). The GSSI’s SIR-30 GPR with 200MHz antenna is presented in (d), which is adopted in our experiments to obtain the B-scan image at each detection line. (e) and (f) show the utilized GPR host and antenna.

When conducting the coordinate system, we chose the direction of the detection line to make it as perpendicular as possible to the direction of the cables on the existing pipeline map. The dimensions of the first and second detection areas are both 20m in length (X axis) and 10m in width (Y axis). The size of the third detection area is 20m long (X axis) and 15m wide (Y axis). In these three areas, detection lines are conducted parallel to each other and also parallel to the Y axis every 2m. For the hyperparameters, $\beta$ is set to be 1, and $\theta_y = 0.3, \theta_z = 0.1$, since the error of the utilized GPR’s supporting positioning equipment could be controlled within 0.3m, while the depths of all detected cables are less than 1m, and the change of the depth of each cable is also less than 0.1m, acknowledged from the existing pipeline map.

B. Experimental Results

The obtained GPR B-scan images are interpreted by the proposed GPR B-scan image interpreting model, and detected points on each detection line are obtained. Due to the limitation of the length of this paper, these GPR images could not be fully demonstrated. Fig. 7 illustrates the process of interpreting a GPR B-scan image, and part of obtained images with processing results are shown in Fig. 8.

![Figure 7](image3.png)

![Figure 8](image4.png)

Fig. 6. Some excavated cables near our experimental areas with probes and trees around. The cables are shown as the white lines. The two figures also demonstrate that the location of buried cables could not be accurately described by straight line segments.

In our experiments, the depths of buried cables are less than 1m with changes less than 0.1m, while the movements of cables are not straightforward. Thus the $XOY$ plane of the established coordinate system is adopted to visualize the cable fitting results as Fig. 9, that is, the same perspective as the satellite map from top to bottom.

The accuracy of the proposed model could be measured by computing the average error $E$ of the depth and position, which could be calculated as

$$E = \frac{1}{m} \sum_{k=1}^{m} |\text{calculated\_value}_k - \text{measured\_value}_k|,$$

where $m$ is the number of data points.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
where the calculated_value indicates the position and depth obtained by the proposed method on $Y$ and $Z$ axes. The measured_value indicates the actual position and depth (values on $Y$ and $Z$ axes) of the points on the buried cable at established detection lines, and some other points randomly selected and measured on the cable. $m$ indicates the number of these points. The actual position and depth of these points are obtained by manually adjusting and fitting the hyperbolas on the GPR images, and we also carried out precise excavation according to the results of our method to confirm the exact depth and location of some points on the cable, without affecting the public facilities. Take the position error on $XOY$ plane as an example. As Fig. 10 shows, when a cable is excavated at a certain position (the blue point in Fig. 10), the $X$ coordinate of the position, that is, $x_k$, will be confirmed in the established coordinate system. The way to confirm $x_k$ is to draw a vertical line from this point to the $X$ axis, and obtain the coordinate of the intersection $(x_k, 0)$ by measuring the distance from $(x_k, 0)$ to the origin $O$. In actual operation, the four vertices of the rectangular detection area ($A$, $B$, $C$, and $O$ in Fig. 10) are marked, so as to ensure the accuracy of vertical line and distance measurement. The coordinate of this excavated position on the $Y$ axis could then be obtained by measuring the distance from $(x_k, 0)$ to the excavated position (the blue point), and is recorded as measured_value_k. The proposed cable fitting method could infer a cable position on $Y$ axis given $x_k$, which is presented as calculated_value_k illustrated as the red point in Fig. 10. Thus the error $E_k$ at $x_k$ could be measured by $E_k = |\text{calculated_value}_k - \text{measured_value}_k|$. The mean value of all errors of a cable could then be calculated by Equation (21).

Experimental results are presented in Table I. The errors of depth and position are within 7cm and 12cm, respectively. In addition, the actual positions of all the points in the experiments are within the obtained confidence intervals. In real-world applications, the obtained interval could provide early warning for excavation work, and it could also reduce the detection range for precise detection between detection lines.
C. Analysis and Comparative Work

In our method, the depth error mainly comes from the process of interpreting GPR B-scan images, especially the hyperbolic fitting. We compared the proposed fitting algorithm with the Restricted Algebraic-Distance-based Fitting algorithm (RADF) [25] and Algebraic-distance-based Fitting Algorithm (ADF) [21], and Fig. 11 shows part of the fitting results. The specific average errors of the proposed fitting algorithm, ADF, and RADF are presented in Table II.

![Fig. 11. (a) and (b) show two examples of the fitting results obtained by the proposed hyperbolic fitting algorithm, RADF, and ADF.](image)

| Area  | The average error of depth (cm) |
|-------|--------------------------------|
|       | Detection line depth | Randomly selected points depth | Altogether depth |
| 1st   | 4.12 | 9.42 | 5.22 | 10.51 |
| 2nd   | 5.92 | 12.52 | 4.81 | 9.88 |
| 3rd   | 5.66 | 11.23 | 7.11 | 9.81 |

TABLE II
END OF THE PROPOSED HYPERBOLIC FITTING ALGORITHM, RADF, AND ADF

When applying ADF, the obtained cable’s depth is greater than the actual value in our experiments. Since the hyperbola has two branches (the upper and lower branches), and some points will be fitted to the upper branch of the hyperbola if no restrictions are imposed. RADF improves accuracy by restricting the center of the hyperbola above the point set on the basis of ADF. However, when detecting objects with a large difference in permittivity with the surrounding medium, such as buried cables, the response signal could be strong, and the obtained point set could be dense. Intuitively, the upper part of the hyperbolic point set is relatively thick. In this case, the attraction of the upper half of the hyperbola to the points could not be completely eliminated through RADF. On the basis of RADF, the proposed fitting algorithm further moves the lower part of the hyperbola to the center of the point set, thereby improving the fitting accuracy. And in our experiments, the numbers of iterations of Gauss-Newton iteration in the proposed hyperbolic fitting algorithm are all less than 10, which verifies the robustness and appropriateness of RADF as the initialization, and also guarantees the efficiency of our model in real-world applications.

In real-world applications, the positioning errors could not be ignored. Satellite-based positioning signals could be blocked [32] in the case of tall buildings and trees. The positioning accuracy of an odometer would degrade when a rough or slippery ground is measured [33]. The proposed cable fitting algorithm could handle detected points with positioning error, and we evaluated the proposed method with the Three-Dimensional Spline Interpolation (TDSI) algorithm [3] and the Marching-Cross-Sections (MCS) algorithm [35], which could both obtain the cable’s location from the detected points at detection lines. Detection lines with intervals of 1m, 2m, and 3m are conducted, and the errors of the three methods are presented in Table III. Due to the limitation of the length of this paper, the results of the conducted experiments could not be fully presented here. The details of some experimental results in different detection line intervals are visualized in Fig. 12.

It could be observed from the results that the error obtained by TDSI is the largest among the three methods when the interval is 2m. This is since TDSI directly utilizes an interpolation method to connect the various acquired cable positions, regardless of the presence of positioning errors. The main concern of MCS is that connecting the points on the cable by straight line segments would cause errors in the cable’s location between connected points. When the distance between the detection lines is 3m, the errors of the three methods are all larger than that with 2m intervals. In particular, the error of MCS changes the most. When the distance between the detection lines becomes larger, the length of the straight line segments obtained by MCS gets longer. Thus the details of the cable’s location between two connected points could not be well described by MCS. When the interval of detection lines is 1m, the errors of MCS and the proposed cable fitting algorithm are reduced, while the error of TDSI becomes larger, since MCS and the proposed cable fitting algorithm consider the existence of noise. When applying these two methods, the greater the density of detected points , the better the denoising effect would obtain and the more accurate the fitting result would become. TDSI does not deal with noises, and when the detected points become dense, the curve obtained by interpolation will get unsmooth with frequent changes of directions. With intervals of 1, 2, and 3m, the accuracy of the proposed algorithm is stable, since both position and depth errors are taken into account and the location of cables is described by curved segments. Moreover, the measured actual positions on the cable are all within the obtained confidence interval.

When the distance between the detection lines is 3m, the errors of the three methods are all larger than that with 2m intervals. In particular, the error of MCS changes the most.


TABLE III
ERRORS OF THE PROPOSED CABLE FITTING ALGORITHM, TDSI, AND MCS UNDER DIFFERENT DETECTION LINE INTERVALS

| Area | The proposed algorithm 1m interval of detection lines | TDSI | MCS | The proposed algorithm 2m interval of detection lines | TDSI | MCS | The proposed algorithm 3m interval of detection lines | TDSI | MCS |
|------|------------------------------------------------------|------|-----|------------------------------------------------------|------|-----|------------------------------------------------------|------|-----|
| 1    | 7.64                                                 | 21.46| 7.65 | 10.14                                                | 17.26| 12.39 | 21.14                                                | 22.18| 29.29|
| 2    | 7.16                                                 | 21.01| 8.21 | 10.76                                                | 14.19| 12.21 | 19.76                                                | 24.46| 31.29|
| 3    | 6.88                                                 | 20.91| 7.02 | 10.28                                                | 15.91| 14.96 | 22.28                                                | 21.51| 29.13|

Fig. 12. These three figures show some details of the results obtained by the proposed cable fitting algorithm, TDSI, and MCS in different detection line intervals. As the spacing between detection lines decreases (e.g. to 1m), the results of MCS and the proposed algorithm will gradually approach the results of the proposed algorithm. The proposed algorithm obtains the results closest to the actual situation under different intervals between detection lines.

When the distance between the detection lines becomes larger, the length of the straight line segments obtained by MCS gets longer. Thus the details of the cable’s location between two connected points could not be well described by MCS. When the interval of detection lines is 1m, the errors of MCS and the proposed cable fitting algorithm are reduced, while the error of TDSI becomes larger, since MCS and the proposed cable fitting algorithm consider the existence of noise. When applying these two methods, the greater the density of detected points are, the better the denoising effect and the more accurate the fitting result become. TDSI dose not deal with noise, and when the detected points become dense, the curve obtained by interpolation will get unsmooth with frequently changes of directions. With intervals of 1, 2 and 3m, the accuracy of the proposed algorithm is stable.

V. CONCLUSION
In this paper, a cable locating method based on GPR and Gaussian process regression is proposed. The coordinate system of the detected area is firstly conducted, and the input and output of locating the buried cables are determined. Parallel detection lines are then established, along which the GPR is moved to obtain the B-scan images. After that, hyperbolic shapes on these obtained images are identified and fitted, thus the positions and depths of some points on the cables could be roughly derived. Finally, these points are clustered and fitted to infer the most likely location of the buried cables. Furthermore, the confidence intervals of cables are also obtained by the proposed method, and in the conducted experiments, the actual positions of the cables we inspected are all within the intervals. In real-world applications, the obtained intervals could be banned from excavation to ensure that the buried cables are not damaged.

VI. ACKNOWLEDGEMENT
This work is sponsored by National Key Research and Development Program of China (No. 2021ZD0111700), National Nature Science Foundation of China( No. 62137002 and 62176245), Key Research and Development Program of Anhui Province(No. 202104a05020011), Key Science and Technology Special Project of Anhui Province (No. 202103a07020002), and the Fundamental Research Funds for the Central Universities.

REFERENCES
[1] N. Metje, P. Atkins, M. Brennan, D. Chapman, H. Lim, J. Machell, J. Muggleton, S. Pennock, J. Ratcliffe, M. Redfern et al., “Mapping the underworld-state-of-the-art review,” Tunnelling and underground space technology, vol. 22, no. 5-6, pp. 568–586, 2007.
[2] N. Yazim, R. Shauri, and N. Buniyamin, “Automated mapping for underground pipelines: An overview,” in 2014 2nd International Conference on Electrical, Electronics and System Engineering. IEEE, 2014, pp. 77–82.
[50] H. Chen, P. Tino, and X. Yao, “Probabilistic classification vector machines,” *IEEE Transactions on Neural Networks*, vol. 20, no. 6, pp. 901–914, 2009.

[51] H. Chen, P. Tino, and X. Yao, “Predictive ensemble pruning by expectation propagation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 7, pp. 999–1013, 2009.

[52] B. Jiang, H. Chen, B. Yuan, and X. Yao, “Scalable graph-based semi-supervised learning through sparse bayesian model,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2758–2771, 2017.

[53] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. Massachusetts Institute of Technology Publishing, 2006, vol. 2, no. 3.

Huanhuan Chen (M’09–SM’16) received the B.Sc. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2004, and the Ph.D. degree in computer science from the University of Birmingham, Birmingham, U.K., in 2008. He is currently a Full Professor with the School of Computer Science and Technology, USTC. His current research interests include neural networks, Bayesian inference, and evolutionary computation. Dr. Chen received the 2015 International Neural Network Society Young Investigator Award, the 2012 IEEE Computational Intelligence Society Outstanding Ph.D. Dissertation Award, the IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award (bestowed in 2011 and only one paper in 2009), and the 2009 British Computer Society Distinguished Dissertations Award. He is now associate editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEM, and the IEEE TRANSITION ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE.

Xiren Zhou received the B.Sc. degree from Shan-dong University, Jinan, China, in 2014, and the Ph.D. degree in computer science from the University of Science and Technology of China, Hefei, China, in 2019. He is currently an associate researcher with the School of Computer Science and Technology, USTC. His research interests include machine learning, ground-penetrating radar, and multi-sensor data fusion.

Qiuju Chen is now an associate research fellow of University of science and technology of China. She received the Ph.D. degree with National University of Defense Technology in 2016. She received the B.Sc. degree in School of information science and technology, University of science and technology of China in 2004. Her research interests include relation extraction and machine learning.

Shengfei Lyu is now a research fellow of Nanyang Technological University (NTU). He received the Ph.D. degree with the School of Computer Science and Technology, University of Science and Technology of China (USTC) in 2020. He received the B.Sc. degree in Information Management and Information Systems from Hefei University of Technology (HFUT) in 2015. His research interests include relation extraction, data mining and machine learning.