Probabilistic Conic Mixture Model and its Applications to Mining Spatial Ground Penetrating Radar Data

Huanhuan Chen* Anthony G Cohn†

Abstract

This paper proposes a probabilistic conic mixture model based on a classification expectation maximization algorithm and applies this algorithm to Ground Penetrating Radar (GPR) spatial data interpretation. Previous work tackling this problem using Hough transform or neural networks for identifying GPR hyperbolae are unsuitable for on-site applications owing to their computational demands and the difficulties of getting sufficient appropriate training data for neural network based approaches. By incorporating a swift conic fitting algorithm into the probabilistic mixture model, the proposed algorithm can identify the hyperbolae in GPR data in real time and further calculate the depth and the size of the buried utility pipes. The number of the hyperbolae can be determined by conducting model selection using a Bayesian information criterion. The experimental results on both the synthetic/simulated and real GPR data show the effectiveness of this algorithm.

1 Introduction

The fitting of primitive models to image data is a basic task in pattern recognition and spatial data mining and also is an important technique for many industrial applications. There are several conic fitting algorithms in the literature [4, 12, 19, 18].

However, most of these algorithms can only identify one conic in each image data and most are sensitive to outliers. An online example of one conic fitting algorithm is published at the following address `http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/PILU1/demo.html`. It is easy to verify that it suffers from the above two shortcomings. However, we also note that this algorithm runs in real-time, even implemented as Java Applet.

To address these two problems and to ensure a fast run time, this paper extends this algorithm using a probabilistic conic mixture model and applies the proposed algorithm to an important application area, Ground Penetrating Radar (GPR) data interpretation.

Ground Penetrating Radar has been widely used as a non-destructive tool for the investigation of the shallow subsurface, and is particularly useful in the detection and mapping of subsurface utilities and other solid objects [9]. However, GPR displays are not easily interpreted and only experts can extract significant information from GPR images to make a reliable report after the inspection.

The patterns appearing in the B-scans [5] of GPR data have shapes determined by the propagation of short pulses into a medium with certain electrical properties. Typically, we can observe hyperbolic curves or linear segments in the GPR image: The first are due to objects with cross-section size of the order of the radar pulse wavelength; the second stem from planar interfaces between layers with different electrical impedances.

As GPR is becoming more and more popular as a shallow subsurface mapping tool, the volume of raw data that need to be analyzed and interpreted is causing more of a challenge. There is a growing demand for automated subsurface mapping techniques that are both robust and rapid. This paper provides such a system.

The current tools that have been developed to aid in GPR data interpretation are generally computationally expensive, such as Hough Transform [16] or neural network based algorithms [3], and inadequate for on-site applications.

By extending a swift conic fitting algorithm in this mixture model, the proposed algorithm can be operated in real time. Other benefits of the proposed algorithm include relative robustness to noise compared with the previous conic algorithms and automatic determination of the number of hyperbolae by a Bayesian information criterion.

The remaining part of this paper is organized as follows. Section 2 will present some relevant works and the algorithm description is described in Section 3. The experimental results are reported in Section 4. Finally, conclusions are drawn in Section 5.

*School of Computing, University of Leeds, Leeds, UK, LS2 9JT. email: H.H.Chen@Leeds.ac.uk

†School of Computing, University of Leeds, Leeds, UK, LS2 9JT. email: A.G.Cohn@Leeds.ac.uk

2 Background

In the literature, there are several published works dealing with the automatic detection of patterns associated with buried objects in GPR data. These algorithms can be grouped into three main categories: 1) Hough transform based methods, 2) machine learning based methods and 3) clustering based algorithms.

Hough transform [16] is a feature extraction technique used in image analysis to find imperfect instances of objects within a certain class of shapes by a voting procedure in a parameter space. The classical Hough transform was concerned with the identification of lines in the image, but later the Hough transform has been extended to identifying positions of arbitrary shapes, most commonly circles or ellipses. Hough transform based methods can identify the four parameters related to the hyperbola, which facilitates subsequent estimation of the pipe size and depth of the buried assets [23, 6]. However, this method often needs to run hundreds of Hough transforms with different combinations of hyperbola parameters (a, b) to search the best fit hyperbola shape and this usually cannot be deployed in real-time applications. Another problem with this kind of algorithm is how to specify a suitable threshold for the number of votes to determine the number of hyperbolae in the image.

There is some work that uses machine learning methods to estimate the size and the depth of the buried pipes. However, with different mediums, soil types, materials of the pipes, the reflected patterns in GPR data are different. In the real-world setting, it is very difficult to acquire the training data for different settings. For example, Pasolli et al. only use simulated data to train the neural networks [17] and this method greatly limits the practical applications.

Some work has been done to use a clustering approach to identify the hyperbolae. In [8], the authors applied a wavelet-based procedure to reduce noise and to enhance signatures in GPR images and then used a fuzzy clustering approach to identify hyperbolae. However, this kind of method will not reveal the hyperbola parameters (a, b) and cannot estimate these parameters related to the buried assets using the geometric model.

In order to address the above problems, this paper proposes a probabilistic hyperbola mixture model. In this model, the feature noise around the hyperbolae and the background noise are both considered. The model is based on a classification expectation maximization (CEM) algorithm [7]. Since it is fast, the algorithm can be deployed in real-time applications.

This algorithm can also be trivially extended to identification of other conic mixtures, such as ellipses and parabolas, thus extending the applicability of the proposed algorithm to many data interpretation scenarios.

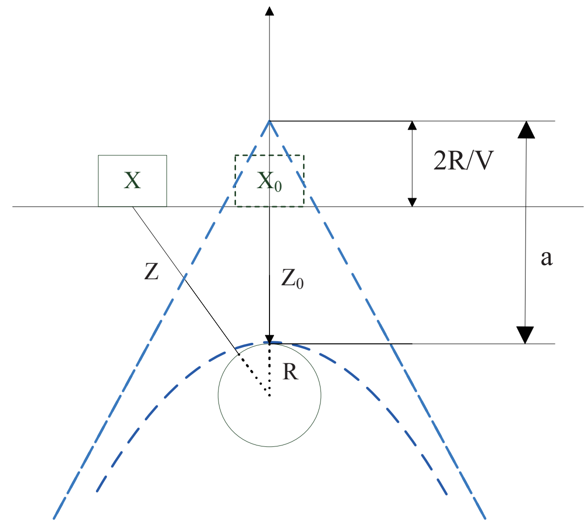


Figure 1: The GPR Geometric Model

3 Probabilistic Conic Mixture Model

In this section, we will present some related knowledge on GPR modeling, a conic fitting algorithm and the probabilistic conic mixture model. In the following subsections, we will present the GPR model description, conic identification algorithm, the probabilistic model, the classification EM algorithm and model selection method using a Bayesian information criterion.

3.1 GPR Model Description The hyperbolic signatures in GPR data are often formulated as a geometric model [22], which is shown in Figure 1. The relation between the two-way travel time t, the horizontal position x and the velocity of propagation v can be expressed by

$$\left(\frac{t + \frac{2R}{v}}{t_0 + \frac{2R}{v}}\right)^2 - \left(\frac{(x - x_0)}{\frac{v}{2}t_0 + R}\right)^2 = 1, \tag{3.1}$$

where (x_0, t_0) are the coordinates of the target, $z = \frac{v}{2}$ and $z_0 = \frac{v_0}{2}$. Equation (3.1) is an equation of a hyperbola centered around $(x_0, \frac{-2R}{v})$.

Relating Equation (3.1) with a general hyperbola,

$$\frac{(y - y_0)^2}{a^2} - \frac{(x - x_0)^2}{b^2} = 1, \tag{3.2}$$

and with some simple derivations, the following relation can be obtained:

$$(3.3)\qquad a = t_0 + \frac{2R}{v},$$

$$(3.4)\qquad b = \frac{v}{2}(t_0 + \frac{2R}{v}).$$

If the parameters related to the hyperbola (a, b) can be found, the depth and the radius can be obtained by the following equations:

$$(3.5)\qquad R = \frac{b(a - t_0)}{a},$$

$$(3.6)\qquad depth = \frac{vt_0}{2} = \frac{bt_0}{a}.$$

This model assumes that a long cylinder is buried in a homogenous medium and the movement of the GPR antenna is perpendicular to the cylinder.

Since most of the pipes are long and linear, in practice, the operator of GPR machine always operates in a perpendicular direction to the assumed direction of the cylinder unless it is suspected that there are T-junctions or the pipes change the direction[1]. The other assumption for the homogenous medium can be satisfied if these pipes are located in the shallow subsurface.

3.2 Hyperbola Fitting Algorithm In this section, we will introduce the algorithm for hyperbola fitting. This single hyperbola fitting algorithm is based on a minor revision of the work [12].

The conic fitting problem can be formulated as an implicit second order polynomial constrained least squares problems.

$$F(\mathbf{A}, \mathbf{x}) = \mathbf{A} \cdot \mathbf{x} = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

where $\mathbf{A} = [A, B, C, D, E, F]^T$ and $\mathbf{x} = [x^2, xy, y^2, x, y, 1]^T$. $F(\mathbf{A}, x_i)$ is called the "algebraic distance" of a point (x_i, y_i) to the conic $F(\mathbf{A}, \mathbf{x}) = 0$.

The shape of the conic function is determined by

$$(3.7)\qquad B^2 - 4AC \begin{cases} > 0 & \text{Hyperbola} \\ = 0 & \text{Parabola} \\ < 0 & \text{Ellipse} \end{cases}.$$

The fitting of a general conic may be approached [15] by minimizing the sum of squared algebraic distances of the curve to the data points. In order to avoid the trivial solution, the parameter vector is often constrained in some way. Many of the published algorithms differ only in the form of constraint applied to these parameters.

Fitzgibbon et al. [12] utilize a constraint on $4AC - B^2 = 1$ for ellipse fitting. In their paper, the Lagrange multiplier and eigen-decomposition are employed to obtain a direct solution, which avoids the iterative optimization and thus performs very fast.

In our paper, we change the constraint to $B^2 - 4AC = 1$ and it will act a base hyperbolic fitting algorithm in the probabilistic mixture model. Based on the formulation in Equation (3.7), our proposed algorithm can fit other conic functions and the combinations of different conic functions, such as elliptic and hyperbolic mixture model.

3.3 Probabilistic Model In practical applications, GPR images are often contaminated with noise. Although various kinds of pre-processing techniques have been proposed to reduce the noise level, it is impossible to guarantee that the processed GPR data is free from noise. In order to take noisy spatial points into consideration, we model two kinds of spatial noise in the proposed algorithm. These two kinds of noise is background noise, in the form of observed points which is not part of the hyperbola and feature noise, which is the deviation of the observed hyperbolic points.

Suppose that X is a set of observation points, and M is a partition consisting of hyperbolae, $M_0, M_1, \cdots, M_K$, where partition M_k contains N_k points. The background noise is denoted by M_0.

In the proposed model, we assume that the background noise is uniformly distributed over the region of the image, which is equivalent to Poisson background noise, and the hyperbolic points are distributed uniformly along the true underlying hyperbola; that is, their algebraic distances[2] follow a normal distribution, with mean zero and variance σ_j^2.

The resulting model becomes a hyperbolic mixture model with the mixing probability π_k $(0 < \pi_k < 1, k = 0, 1, \cdots, K$, and $\sum_{k=0}^{K} \pi_k = 1)$. Then the likelihood can be expressed by

$$L(X|\pi, \sigma) = \prod_{i=1}^{N} L(x_i|\pi, \sigma),$$

where $L(x_i|\pi, \sigma) = \sum_{k=0}^{K} \pi_k L(x_i|\pi_k, \sigma_k, x_i \in M_k)$ and the

[1] Utility map records, which although notoriously inaccurate, at least in the UK, generally give the rough direction of the line of the buried apparatus (which is typically along the line of the road).

[2] We use algebraic distance instead of other robust distance metrics, such as orthogonal distance [2], to fast fit the hyperbola and to be suitable for on-site applications. The robustness to noise will be enhanced by the probabilistic model.

$$L(x_i|\pi_k, \sigma_k, x_i \in M_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{\|f_k(x_i)\|^2}{2\sigma_k^2}\right),$$

where $f_k(x_i)$ is the algebraic distance from the point (x_i, y_i) to the kth hyperbola.

For background noise, the likelihood can be expressed by

$$L(x_i|\pi_0, \sigma_0, x_i \in M_0) = \frac{1}{Area},$$

where $Area$ is the area of the image.

3.4 Classification Expectation Maximization Algorithm The classification expectation maximization (CEM) algorithm is a classification version of the well-known EM algorithm [10]: it incorporates a classification step between the E-step and the M-step of the EM algorithm using a maximum a posteriori (MAP) principle. In the following, the procedure of the algorithm is presented with CEM.

Firstly, we provide the number of hyperbolae in the GPR data and start with an initial partition using the k-means algorithm.

1. Begin with an initial partition.

2. (M-step) With the configuration of the current partitions, fit a hyperbola to each partition and then compute the maximum likelihood estimates (π_k^m, σ_k^2) for $k = 1, \cdots, K$.

$$\pi_k^m = \frac{\#\pi_k^{m-1}}{N},$$

and

$$\sigma_k^2 = \frac{1}{\#\pi_k^{m-1}} \sum_{x_i \in \pi_k^{m-1}} (f_k(x_i) - \bar{f}_k(x_i))^2,$$

where $f_k(x_i)$ is the algebraic distance from the point (x_i, y_i) to the kth hyperbola. π_0^m can be estimated by $\frac{\#\pi_0^{m-1}}{N}$.

3. (E-step) Based on the current hyperbolae and parameter estimates, calculate the likelihood of each point being in each partition.

$$t_k^m(x_i) = \frac{\pi_k^m L_k(x_i)}{\sum_{k=0}^K \pi_k^m L_k(x_i)},$$

where $L_k(x_i) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{\|f_k(x_i)\|^2}{2\sigma_k^2}\right)$, $k = 1, \cdots, K$ and $L_0(x_i) = \frac{1}{Area}$.

4. (Classification step) Assign each point x_i to the partition which provides the maximum posterior probability $t_k^m(x_i)$, $0 \le k \le K$, (if the maximum posterior probability is not unique, we choose the partition with the smallest index).

5. Check for convergence; end or return to Step 2.

After we calculated the probability of each point being in each partition, we assign each point into the partition for which it has the highest likelihood. Note that at the end of each iteration, the likelihood of the model will be calculated. Since the classification expectation maximization iterations sometimes decrease the likelihood, the process is executed for a predetermined number of iterations, and we choose the model with the highest overall likelihood as the final result.

3.5 Model Selection by Bayesian Information Criterion As with other mixture models, the hyperbolic mixture model needs to specify the number of hyperbolae at the beginning. The usual strategy is to search a range for the number of hyperbolae k and select the best one based on proper model selection methods.

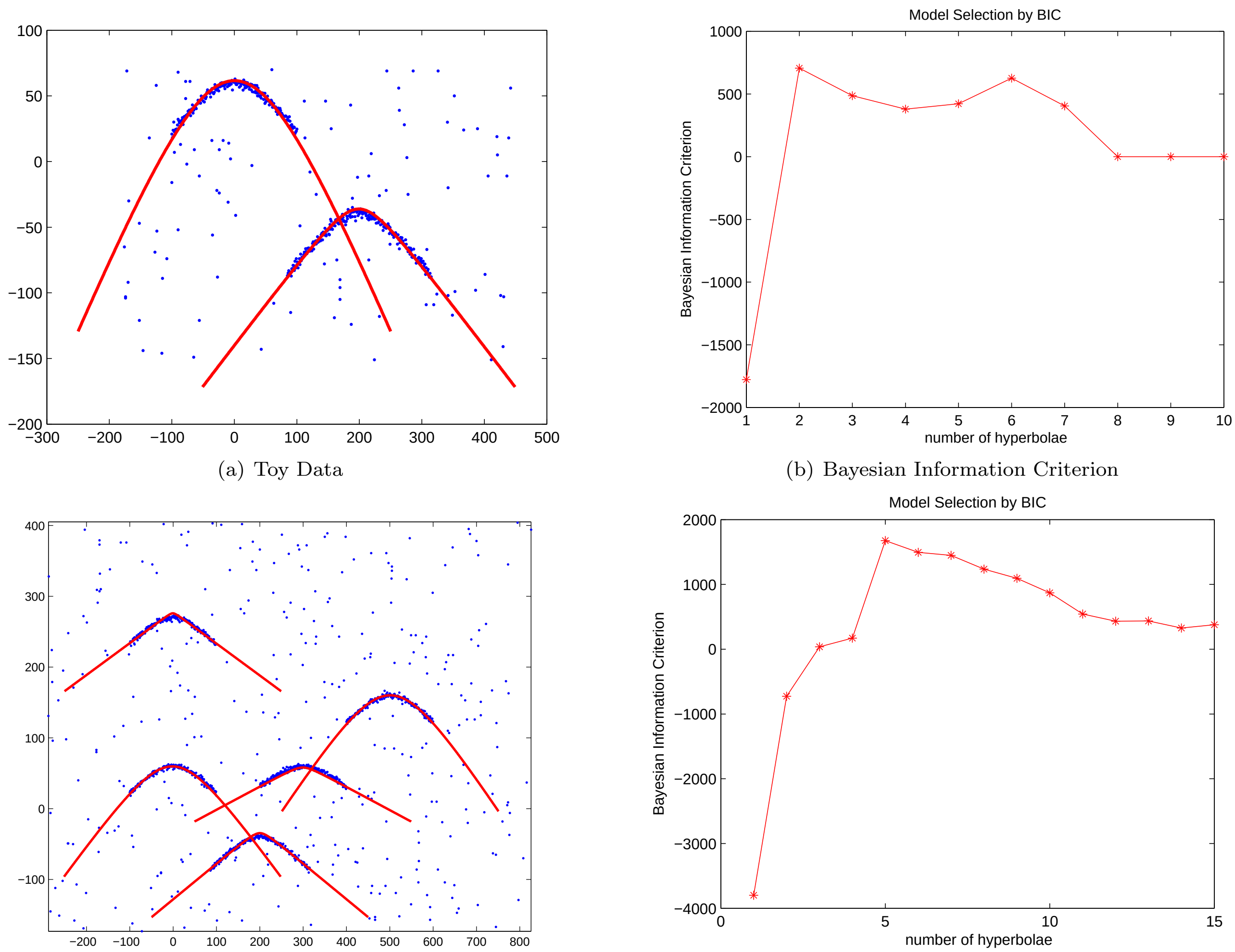
In this paper, we propose to use a Bayesian information criterion (BIC) [21] for model selection among a class of parametric models with different numbers of parameters. The model selection based on BIC can be seen as a form of regularization since it is possible to increase the likelihood by adding additional parameters in the maximum likelihood estimation, which may lead to overfitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model.

Model selection based on BIC provides (asymptotically) consistent estimators of the probability distribution giving a data set [20]. This approach works well in practice for mixture models and other model-based clustering problems [13, 20]. The BIC for a model with K hyperbolae and background noise is defined by:

$$BIC = 2\log(L) - M\log(N),$$

where $M = K(DF + 2) + K + 1$ is the number of parameters, DF is the degrees of freedom used in fitting a hyperbola and there are four degrees of freedom in each hyperbola, i.e. $DF = 4$. The number of hyperbolae is K; for each hyperbola, we need to estimate σ_j, and we fit a hyperbola using four degrees of freedom. There are K parameters associated with the mixing proportions[3] and one more parameter is used for

[3] Although there are $M + 1$ mixing coefficients, the constraint $\sum_{k=0}^K \pi_k = 1$ reduces one degree of freedom.



(a) Toy Data (b) Bayesian Information Criterion

(c) Synthetic Data (d) Bayesian Information Criterion

Figure 2: Hyperbola Identification and Model Selection for Synthetic Data

image area estimation. The larger the BIC, the more the model is favoured by the data.

As with other mixture models, our algorithm is still sensitive to the initializations. In this paper, we utilize a k-means algorithm to initialize the partitions. In the experimental sections, we emphasize that the number of data points in the primary hyperbola is significantly more than the number of noise points. In order to further reduce the impact of the initializations, the experiments in this paper are run five times by providing different random seeds for the k-means algorithm and the model with the highest BIC value will be selected.

4 Experimental Study

In order to examine the proposed algorithm, this section conducts several experiments on simulated and real GPR images. The precision and the computational cost are also analyzed in this section.

4.1 Synthetic Data In this subsection, two synthetic data with two and five hyperbolae are generated with Gaussian white noise, respectively. These hyperbolae are positioned in different locations and have similar shape to reflect the case with real GPR data. Since in GPR B-scan images, the shape of the hyperbola is only determined by the medium where objects are buried [11] and if we assume that the medium does not change dramatically over a small neighbourhood, then the reflected hyperbolae should have similar shapes.

Figure 2 illustrates the results. From these figures, it can be seen that the proposed algorithm successfully identifies these hyperbolae and ignores the noise points.

To select the most appropriate number of hyperbo-

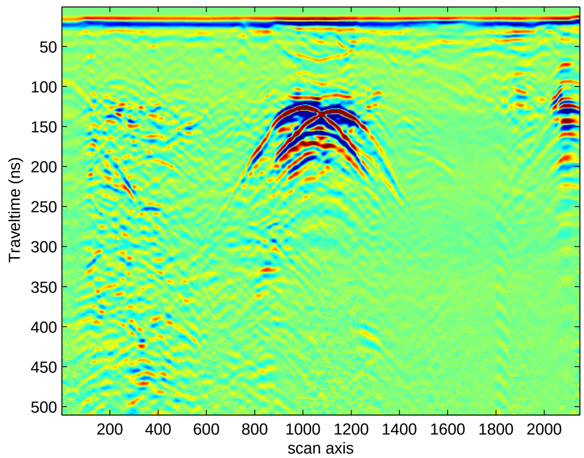


Figure 3: B-scan GPR Data

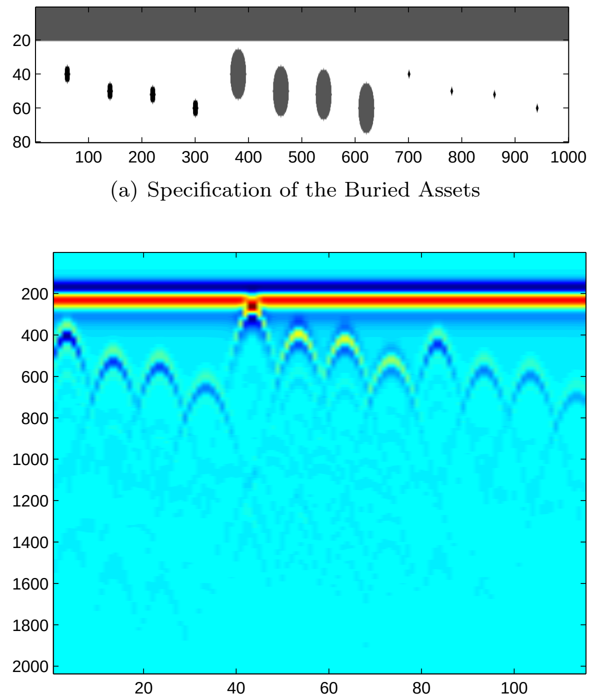
lae for this data set, we run the algorithm with different k, which is the number of hyperbolae in the data set, and record the BIC value. The result confirms that two and five hyperbolae are the appropriate number of hyperbolae for these two data sets, respectively.

4.2 Real GPR Data In this subsection, we utilize a real GPR data set to validate the proposed algorithm. The B-scan image is illustrated in Figure 3. In this figure it can be seen that the data set is challenging since it contains significant noise and has another unwanted secondary hyperbola in the middle of two primary hyperbolae. This unwanted secondary hyperbola is named a "ghost area" in GPR industry and it is generated by the cross-reflection of two nearby buried objects.

In order to process this data, we preprocess to reduce the noise, remove the background to delete the linear reflection of ground (upper part in Figure 3) and reduce the intensity of the GPR data to reduce the impact of the secondary hyperbola. For each pixel in GPR image, the intensity of each non-zero point is ranked and the first quartile, i.e. the median of the data which are less than the overall median, of ranked intensity is selected as the cut off value to prune the noise points. Another method could sample several intensities and use BIC to select the most appropriate one.

After the preprocessing step, the proposed algorithm is applied to this data. Figure 4 illustrates the results. From this figure, it can be seen that the proposed algorithm successfully identifies these hyperbolae. Based on the BIC figure, we also notice that the proposed algorithm with over-estimated k often generates a higher BIC value than the mode with under-estimated k. This is due to the existence of substantial noise.

We will select the model with relatively small k from some candidate models whose BIC values are similar in the following processing[4].



(a) Specification of the Buried Assets

(b) Simulated GPR Data

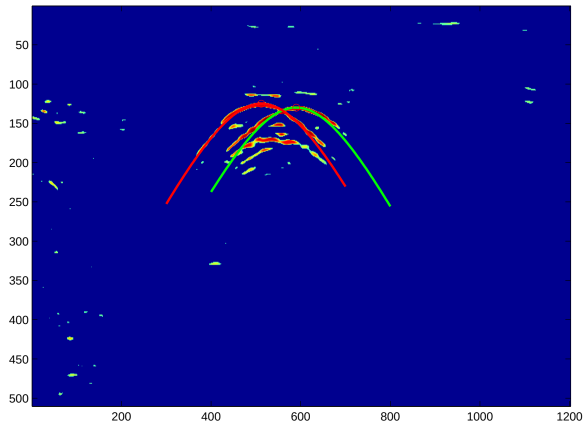
Figure 5: The Buried Assets and the simulated GPR Data

4.3 Simulated GPR Data In practice, at least in the UK, utility records rarely contain depth information (not withstanding its potential usefulness), so evaluating the depth estimate from our algorithm without physical excavation is difficult. Size information is usually present in the statutory records, but can not be completely relied upon for accuracy.

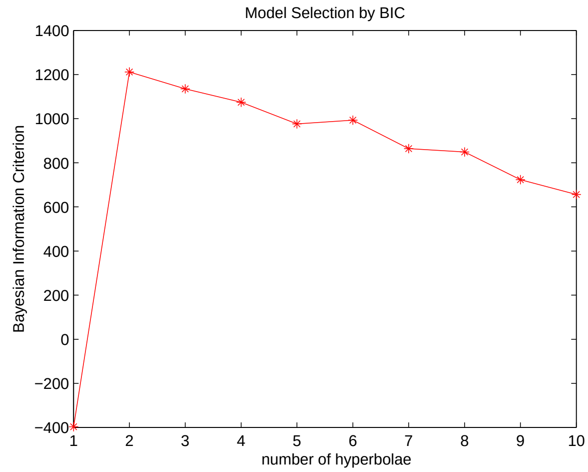
To resolve this problem, in this section we employ simulated GPR data to estimate the accuracy for estimation of the depth and size of buried assets.

The simulated GPR data is generated by means of the electromagnetic simulator GprMax [14]. GprMax

[4] In practice, in the utility sector since, as already noted, buried apparatus is typically linear and relatively long in length, further evidence for the number of hyperbolae/buried objects will come from repeated GPR measurements at regular intervals (typically at least three scans are taken 1m apart along the length of the suspected apparatus. By integrating the evidence over these multiple scans, the estimate on k can be further improved.



(a) Real GPR Data



(b) Bayesian Information Criterion

Figure 4: Hyperbola Identification for GPR Data

was developed on the basis of the Finite-difference time-domain (FDTD) numerical method. It discretizes the Maxwell's equations in both space and time and obtain an approximate solution directly in the time domain through an iterative process. GprMax allows the user to specify different mediums, such as clay, soft sand, concrete, and different sizes of buried objects with varied diameters.

In Figure 5, we show an example to specify the buried assets and the obtained simulated GPR Data. Note that all the buried assets are assumed to be cylinders. In order to validate our algorithm, we have generated one simulated GPR dataset with one hundred of buried pipes with varied sizes in varied mediums. The radii of these pipes range from 4cm to 20cm and the depths range from 40cm to 120cm. Note that in order to obtain better results, we need to generate a relatively higher resolution GPR data to guarantee the number of data points in the primary hyperbola is significantly more than the number of noise points.

In this experiment, one variant of the proposed algorithm, hyperbolic-$k$means, is included for comparison. The hyperbolic-$k$means algorithm obtains these hyperbolae according to the principle of k-means and assigns the points to the hyperbola with the shortest algebraic distance. Since we could not define a likelihood function for the non-probabilistic model, hyperbolic-$k$means, BIC will not be used in hyperbolic-$k$means algorithm and the k will be chosen as the same value as the one in Hyperbolic-mixture, which is optimized by BIC. Since hyperbolic-$k$means does not take the probabilistic model into consideration and it is very sensitive to noise.

Table 1 reports the statistical results of the experiment. The proposed algorithm manages to identify 94 out of 100 hyperbolae and for the identified hyperbolae, the obtained hypotheses on the depth and size are quite accurate. The number of hyperbolae k, selected by BIC, is a little greater (105) than 100. This is because the secondary hyperbolae of some pipes with large diameters, buried in shallow subsurface, have large intensity even after a series of pre-processing steps.

Compared with our algorithm, hyperbolic-$k$means, which does not incorporate the probabilistic model, only identifies 53 hyperbolae out of 100 and the calculated hypotheses are significantly worse than our algorithm in terms of the accuracy. It is also worth mentioning that both algorithms operate almost in real-time (less than 1 second for the GPR dataset). Based on this experiment, the probabilistic conic mixture model achieves satisfactory performance in terms of the accuracy and time.

We use algebraic distance instead of other robust distance metrics, such as orthogonal distance, to fast fit the hyperbola since most of robust distance fitting algorithms have a high computational complexity [1]. For example, the orthogonal distance fitting has to be iteratively implemented and if it was incorporated into the mixture model, it will be difficult to be deployed for on-site applications due to the computational demands. In this method, the robustness to noise is enhanced by the probabilistic model.

5 Conclusions

Previous algorithms for mining GPR data are unsuitable for on-site applications due to

Table 1: The Comparison between three algorithms on hyperbola identification and hypotheses extraction from GPR data. The number of hyperbolae k ranges from 1 to 150 and the actual number of hyperbolae in the data set is 100.

Algorithm	Hyperbola Identify (#)	k selected by BIC	Running Time	Depth Error (%)	R Error (%)
Hyperbolic-mixture	94	105	0.8s	5.4	3.79
Hyperbolic-$k$means	53	105(fixed)	0.3s	17.5	14.3
Hough Transform	89	105(fixed)	226.1s	4.9	4.2

computational complexity or the difficulty of obtaining sufficient appropriate training data for neural network based methods. We have addressed both of these problems in this paper.

In order to develop a novel GPR data mining algorithm, we extend an existing *single* conic fitting algorithm by incorporating a probabilistic conic mixture model and employing the classification expectation maximization algorithm for the final solution.

The proposed algorithm significantly contributes to both theoretical research and practical application areas. From a theoretical point of view, this research extends the existing single conic fitting algorithm to multiple conic fitting algorithm and provides a robust and swift solution compared to the previous conic fitting algorithms. In this paper, we mainly focus on mixtures of hyperbolae although the proposed search can be trivially extended for other conics, such as ellipses and parabola.

For practical applications, the proposed techniques provide an effective and accurate GPR data interpretation tool, which is adequate for on-site applications. This is extremely useful for the advanced multi-channel GPR system that often generates volumes of data in each task. Therefore, this research will potentially play an important role in GPR and the related industries, such as utility detection, infrastructure and transportation industries.

In this paper, a swift conic fitting algorithm, which is sensitive to noise, based on the eigen-decomposition is employed based on the consideration for real-time deployment of the proposed algorithm.

Therefore, the number of data points in the primary hyperbolae should be significantly more than the number of noise points. This then allows preprocessing to reduce noise before our algorithm is applied.

Our method can be seen as a trade-off between computational time and robustness. Our future work will focus on improving the robustness of this probabilistic conic mixture model. As remarked earlier, one other way will be to incorporate evidence from multiple scans along the length of a suspected linear object.

Acknowledgement

This research is supported by an EPSRC grant (EP/F06585X/1): Mapping The Underworld: Multi-sensor Device Creation, Assessment, Protocols.

References

[1] S.J. Ahn. *Least squares orthogonal distance fitting of curves and surfaces in space.* Springer-Verlag New York Inc, 2004.

[2] S.J. Ahn, W. Rauh, H.S. Cho, and H.J. Warnecke. Orthogonal distance fitting of implicit curves and surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 620–638, 2002.

[3] W. Alnuaimy, Y. Huang, M. Nakhkash, M. Fang, V. Nguyen, and A. Eriksen. Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition. *Journal of Applied Geophysics*, 43:157–165, 2000.

[4] F. L. Bookstein. Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, 9(1):56–71, 1979.

[5] C. Bruschini, B. Gros, F. Guerne, P. Y. Pièce, and O. Carmona. Ground penetrating radar and imaging metal detector for antipersonnel mine detection. *Journal of Applied Geophysics*, 40(1-3):59–71, 1998.

[6] L. Capineri, P. Grande, and JAG Temple. Advanced image-processing technique for real-time interpretation of ground-penetrating radar images. *International Journal of Imaging Systems and Technology*, 9(1):51–59, 1998.

[7] G. Celeux and G. Govaert. A Classification EM algorithm for clustering and two stochastic versions. *Computational Statistics & Data Analysis*, 14(3):315–332, 1992.

[8] S. Delbo, P. Gamba, and D. Roccato. A fuzzy shell clustering approach to recognize hyperbolic signatures in subsurface radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1447–1451, 2000.

[9] A. Dell'Acqua, A. Sarti, S. Tubaro, and L. Zanzi. Detection of linear objects in GPR data. *Signal Processing*, 84(4):785–799, 2004.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal Of The Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[11] A. Dolgiy, A. Dolgiy, and V. Zolotarev. Optimal radius estimation for subsurface pipes detected by ground penetrating radar. In *Proceedings 11th International Conference on Ground Penetrating Radar, Columbus, Ohio, USA*, volume 4, 2006.

[12] A. Fitzgibbon, M. Pilu, and R. B. Fisher. Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480, 1999.

[13] A. Gasgupta and A.E. Raftery. Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association*, 93(441), 1998.

[14] A. Giannopoulos. Modelling ground penetrating radar by GprMax. *Construction and Building Materials*, 19(10):755–762, 2005.

[15] R.M. Haralick and L.G. Shapiro. *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1992.

[16] J. Illingworth and J. Kittler. A survey of the hough transform. *Computer vision, graphics, and image processing*, 44(1):87–116, 1988.

[17] E. Pasolli, F. Melgani, and M. Donelli. Automatic analysis of GPR Images: a pattern-recognition approach. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2206–2217, 2009.

[18] M. Pilu, A. Fitzgibbon, and R. Fisher. Ellipse-specific direct least-square fitting. In *Proceedings of International Conference on Image Processing (ICIP'06)*, volume 3, 1996.

[19] J. Porrill. Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter. *Image and Vision Computing*, 8(1):37–41, 1990.

[20] K. Roeder and L. Wasserman. Practical bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, 92:894–902, 1995.

[21] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

[22] S. Shihab and W. Al-Nuaimy. Radius estimation for cylindrical objects detected by ground penetrating radar. *Sensing and Imaging: An International Journal*, 6(2):151–166, 2005.

[23] C. G. Windsor, L. Capineri, and P. Falorni. The estimation of buried pipe diameters by generalized hough transform of radar data. In *Proceedings Progress In Electromagnetics Research Symposium (PIERS)*, pages 22–26, 2005.