# Buried Utility Pipeline Mapping Based on Multiple Spatial Data Sources: A Bayesian Data Fusion Approach

Huanhuan Chen and Anthony G Cohn
University of Leeds

## Background

- ~4 million holes are cut annually into UK streets in order to install or repair buried assets.
- Existing records are notoriously inaccurate and incomplete.
- Failure to identify accurately the location of existing buried assets results in numerous practical problems, costs and dangers for both authorities and road users.
- The *Mapping the Underworld (MTU)* project seeks to develop the means to locate, map in 3-D and record, the position of buried utility service pipes and cables without excavation.


## The VISTA and MTU Projects

- Following on from an initial MTU project, the £2.4 million TSB-funded VISTA project investigated the use of global navigation satellite technology linked to existing asset records to produce visualisations of utilities' underground assets. The project was carried out by the Universities of Leeds and Nottingham in collaboration with more than 20 stakeholders. A key output from the project was a set of techniques to syntactically and semantically integrate records from across the sector; these were validated in several trials across the UK.

- The second phase of MTU aims to build a prototype multi-sensor device with the objective of locating, mapping in 3-D and recording the position of all buried utility service pipes and cables without excavation. – a "body scanner for the street". This multisite project with six universities has over 30 industrial partners. The University of Leeds is responsible for the work package on fusing of the sensor data with the integrated maps (developed under VISTA) which represent the locations of buried assets as recorded by the utilities. The challenge is to reconcile the potentially noisy sensor data with the inaccurate maps from the utilities.

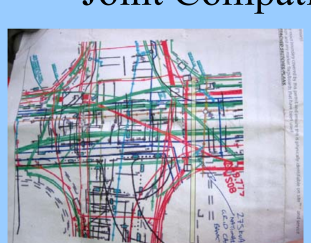
**MTU-2 objectives**

- Create a multi-sensor device that combines complementary technologies for simultaneous surface and in pipe deployment. These sensors including Ground penetrating radar (GPR), Acoustics, passive Magnetic Fields and Low Frequency EM (LFEM) fields.
- Intelligently fuse the outputs of the four sensors to create cross-sections through the ground showing the probability of service locations, taking in to account the expectations from the utility records.

## Utility Pipeline Mapping Based on Bayesian Data Fusion

**Data Sources:**

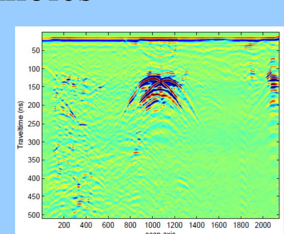
1) Prior: statutory records from VISTA database (inaccurate and incomplete)
   - Given a point of interest (PoI) $(x, y)$, search for nearby pipes in records to find $V = (x, y, \Theta)$, where $\Theta$ is the direction of a pipe passing PoI.
   - Uncertainty is represented by a diagonal matrix $\mathbf{C}_v = \mathrm{diag}(\Delta x, \Delta y, \Delta\Theta)$ where $\Delta x$, $\Delta y$ and $\Delta\Theta$ are the uncertainty variables.
2) Street furniture data (e.g. manhole observations)
   - Consists of a series of manhole locations and estimated asset directions, along with an uncertainty matrix.
3) Ground penetrating radar (GPR) site survey.
   - Hyperbolae in B-scan images represent possible pipes
   - Adjacent B-scans can be used to estimate direction of a pipe segment [Chen &Cohn 2010a]
   - Each observed pipe can be represented by a vector $\boldsymbol{O} = (x, y, \Theta)$ and the uncertainty is represented by a matrix $\mathbf{C}_o$.
4) Pipe linearity assumption $\boldsymbol{L}$: *indirect* data source based on the observations of other pipe locations.
   - Most pipes are approximately linear, this variable provides evidence as to whether to connect pipes segments detected by GPR scans or manhole inspections.
   - Joint Compatibility Branch and Bound (JCBB) used to connect pipes and manholes


Statutory Records in Sheffield City Centre


Manhole observations


*A single-channel GPR*


*Multi-channel GPR*


A typical B-Scan GPR image

### Bayesian Data fusion (BDF) Algorithm:

Suppose there is a variable of interest $t = (x, y, \Theta)$ at a location $(x, y)$.

- This data at $(x, y)$ is related to the variable of interest $\mathbf{t}$ through error terms $\mathbf{E}_o$ and $\mathbf{E}_l$:

$$\begin{aligned} \boldsymbol{O} &= \mathbf{t} + \mathbf{E}_o, \\ \mathbf{L} &= \mathbf{t} + \mathbf{E}_l. \end{aligned}$$

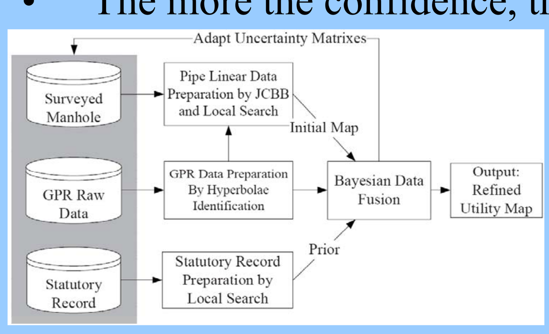
- The posterior probability $p(t \mid O, L)$, given the information O, L at location $(x, y)$ is

$$p(\mathbf{t}|\mathbf{O},\mathbf{L}) = \frac{p(\mathbf{O},\mathbf{L}|\mathbf{t})p(\mathbf{t})}{p(\mathbf{O},\mathbf{L})}.$$

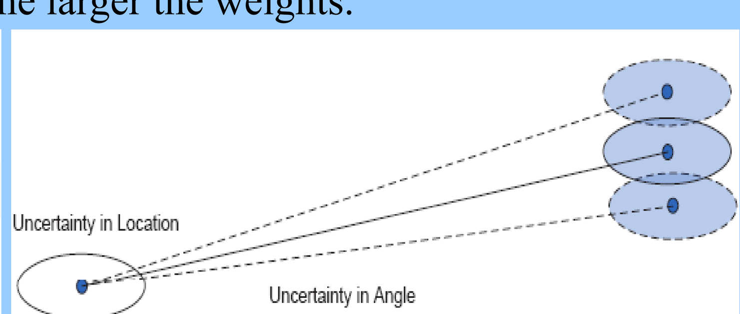
- Assume the mutual independence of data sources, and $p(t \mid O)$ and $p(t \mid L)$ follow mean-zero Gaussian distribution, the posterior probability $p(t \mid O, L)$ is also a Gaussian $N(m, \mathbf{C})$:

$$\begin{aligned} \mathbf{m} &= (\mathbf{C}_o^{-1} + \mathbf{C}_l^{-1} + \mathbf{C}_v^{-1})^{-1}(\mathbf{C}_o^{-1}\mathbf{O} + \mathbf{C}_l^{-1}\mathbf{L} + \mathbf{C}_v^{-1}\mathbf{V}) \\ \mathbf{C} &= (\mathbf{C}_o^{-1} + \mathbf{C}_l^{-1} + \mathbf{C}_v^{-1})^{-1}. \end{aligned}$$

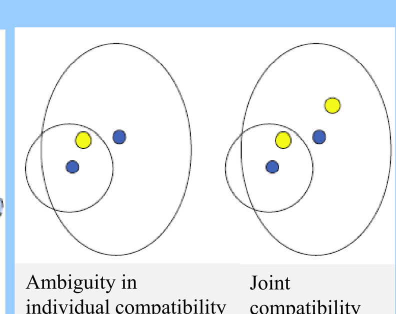
- The Bayesian fusion rule is a weighted average of these three predictions.
- The weights are determined by the confidence value.
- The more the confidence, the larger the weights.



Flowchart of BDF Algorithm



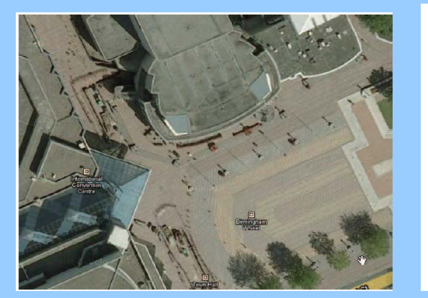
Uncertainty Expansion

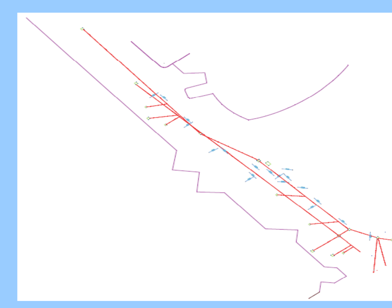


JCBB Algorithm

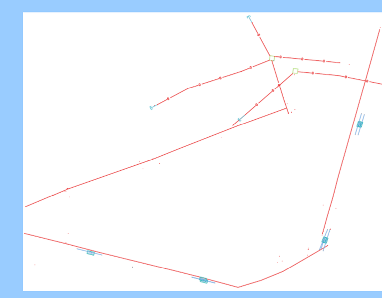
## Experimental Results

- Two real data sets from the UK: each data set consists of an AutoCAD drawing (representing statutory records), a set of GPR scans and the street survey results.

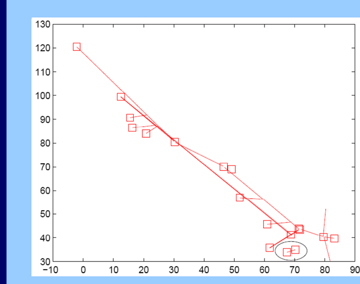

Site 1 Satellite Image

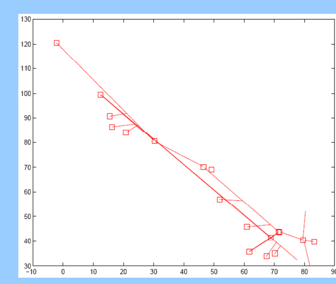

Site 1 Records

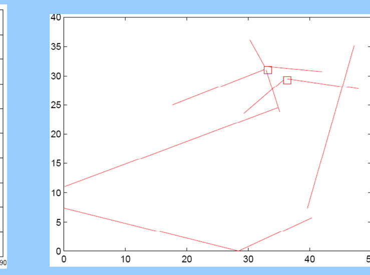

Site 2 Satellite Image


Site 2 Records

- The GPR scan boxes (small box with a line going through the box) are also illustrated in the 2nd and 4th figures. The line going through the box indicates the direction of GPR scan and the GPR machine operates orthogonally to this line.
- Uncertainty: the manhole observations: $\mathbf{C}_o$= diag(0.2,0.2,8) , GPR scans $\mathbf{C}_o$= diag(0.4,0.4,10) , statutory records $\mathbf{C}_v$= diag(1,1,8) , pipe linearity assumption $\mathbf{C}_l$= diag(1.5,1.5,15) .

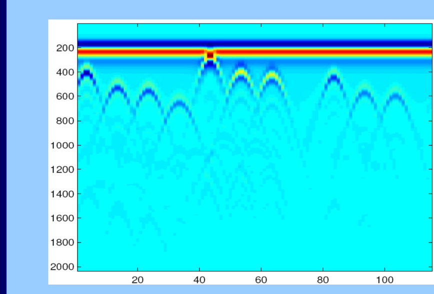

Site 1 without statutory record

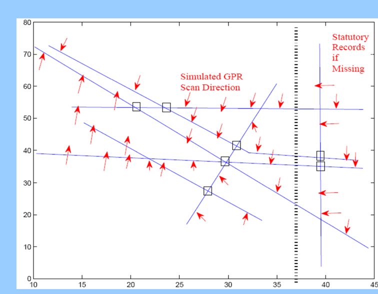

Site 1 with statutory record

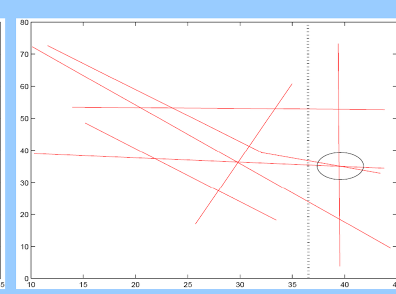

Site 2 with statutory record

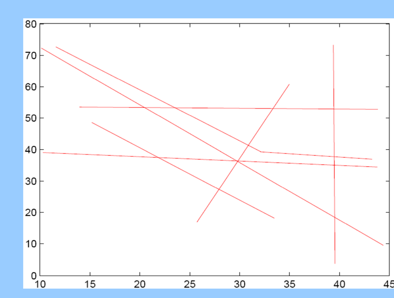
| Data | Area (m×m) | # Pipe segments | # Manhole | #Scans |
|---|---|---|---|---|
| R1 | 100×100 | 19 | 18 | 23 |
| R2 | 100×100 | 10 | 2 | 26 |
| S1 | 100×100 | 22 | 7 | 36 |
| S2 | 100×100 | 38 | 10 | 54 |
| S3 | 100×100 | 44 | 18 | 76 |
| S4 | 100×100 | 61 | 21 | 93 |

Summary information for two real data sets (R1, R2) and four simulated data sets (S1, S2, S3 and S4).


Simulated GPR Result


Simulated Data Set


Mapping without statutory record


Mapping with statutory record

| #error | BDF | BDF\V | BDF\L | BDF\GPR | JCBB(O) | V | Time(s) |
|---|---|---|---|---|---|---|---|
| R1 | 0 | 2 | 1 | 3 | 5 | 1 | 6.6 |
| R2 | 0 | 0 | 0 | 0 | 0 | 0 | 2.9 |
| S1 | 0 | 3 | 2 | 5 | 5 | 2 | 9.3 |
| S2 | 1 | 2 | 1 | 4 | 3 | 2 | 10.2 |
| S3 | 1 | 1 | 1 | 6 | 4 | 1 | 13.7 |
| S4 | 2 | 4 | 3 | 8 | 6 | 3 | 14.6 |

The connection errors of BDF, BDF without statutory records (BDF\V), BDF without pipe linearity assumption (BDF\L), BDF without GPR (BDF\GPR), JCBB on only GPR/manhole survey (JCBBO), and statutory records (V) on two real-world data sets (R1 and R2) and four simulated data sets (S1-S4).

| spatial error | BDF | BDF\V | BDF\L | BDF\GPR | JCBB(O) | V |
|---|---|---|---|---|---|---|
| $E(x,\theta)_{R1}$ | 0.3,3.3 | 0.5,7.2 | 0.4,4.7 | 0.7,5.8 | 0.7,9.4 | 0.8,6.5 |
| $C(x,\theta)_{R1}$ | 0.2,3.6 | 0.3,6.5 | 0.2,4.5 | 0.4,4.6 | 0.3,10.8 | 1,8 |
| $E(x,\theta)_{R2}$ | 0.4,2.8 | 0.4,6.3 | 0.4,4.1 | 0.9,5.4 | 0.8,10.1 | 0.6,6.1 |
| $C(x,\theta)_{R2}$ | 0.2,3.8 | 0.3,7.3 | 0.3,4.9 | 0.6,5.9 | 0.4,12.6 | 1,8 |
| $E(x,\theta)_{S1}$ | 0.3,3.1 | 0.5,7.9 | 0.3,5.0 | 0.9,8.6 | 0.8,10.6 | 0.7,6.9 |
| $C(x,\theta)_{S1}$ | 0.2,3.8 | 0.3,7.1 | 0.3,4.8 | 0.5,7.3 | 0.4,12.2 | 1,8 |
| $E(x,\theta)_{S2}$ | 0.3,3.4 | 0.5,7.9 | 0.5,5.1 | 0.9,9.3 | 0.8,10.7 | 0.7,7.1 |
| $C(x,\theta)_{S2}$ | 0.2,3.8 | 0.3,7.1 | 0.3,4.8 | 0.5,7.6 | 0.4,12.2 | 1,8 |
| $E(x,\theta)_{S3}$ | 0.4,3.6 | 0.8,8.4 | 0.5,5.3 | 0.9,10.2 | 0.9,11.2 | 0.8,7.2 |
| $C(x,\theta)_{S3}$ | 0.2,3.8 | 0.3,7.1 | 0.3,4.8 | 0.6,8.9 | 0.4,12.0 | 1,8 |
| $E(x,\theta)_{S4}$ | 0.4,3.8 | 0.9,8.6 | 0.5,5.6 | 1.0,11.1 | 0.9,11.5 | 0.8,6.8 |
| $C(x,\theta)_{S4}$ | 0.2,3.8 | 0.3,7.1 | 0.3,4.8 | 0.5,8.1 | 0.4,12.1 | 1,8 |

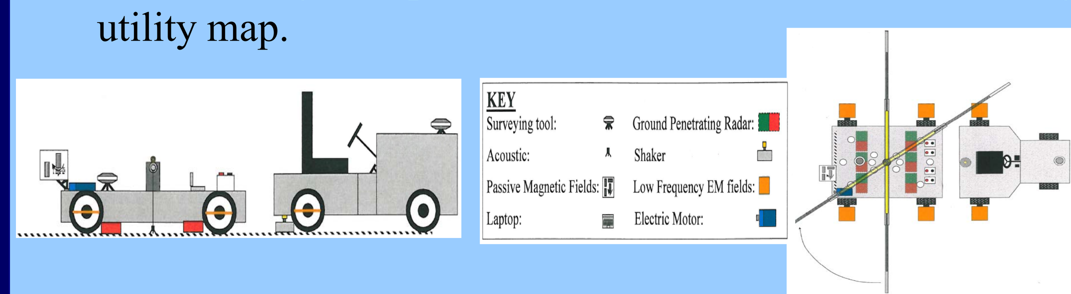
The spatial errors and uncertainty. E(x, Θ) represents the mean spatial distance (in metres) from the real PoI from the estimated PoI and the mean difference (in degrees) of real pipe direction and the estimated direction. C(x, Θ) stands for the uncertainty of these two terms.

- BDF with full data sources outperforms other algorithms.
- JCBB with only GPR/manhole survey is the worst.
- **BDF\V** is equal or inferior to BDF**L**, indicating the statutory records usually contain more information, if presented, than the pipe linearity assumption.
- In general it is very unlikely that only using observations of street furniture (such as manholes) and statutory records will give good results (some buried utilities may well have no such street furniture in the surveyed area). Therefore, BDF\GPR generates inferior performance than BDF with all data sources.
- The computation time are recorded on a 2.4Ghz laptop with 4GB memory on a single core. Clearly, the algorithm can operate in real time as data is gathered (given the push speed of the GPR).

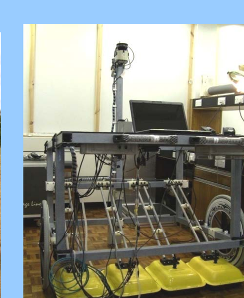
All these results confirm the benefits of inclusion of more data sources and the effectiveness of the BDF algorithm in utility mapping.

## MTU Trolley and AI Subsystem

- The current survey system includes the operator for GPR machine with global positioning system (GPS) for location identification, a surveyor for manhole locations, a GPR machine, statutory records, and the AI subsystem.
- It takes typically takes a three person surveying team a full day to map 150 linear metres over a typical 10m wide highway in a non city centre position with no traffic management issues or other restrictions. Much of the time is not spent in physically obtaining the data but rather in processing and connecting it – exactly the topic of our automated algorithm.
- The AI subsystem provides an automatic and effective way for on-site work to produce a buried utility map.



MTU Trolley with GPR, Sonar, passive magnetic fields and low frequency EM fields.



Our GPR arrays and low frequency EM sensors.

## Summary

- Previous approaches to produce buried utility pipeline maps depend on manual drawing and expert interpretation of GPR scans.
- We believe this paper represents the first attempt to automatically map utility data.
- Algorithms for Bayesian data fusion (BDF) of multiple data sources to connect these manholes and GPR scan locations.
- Comparison of BDF methods with different combinations of data sources.
- The uncertainty of both the location and direction of pipes are considered in the algorithm.

## Future Work

- Incorporate expectations from other sensors, e.g. sonar, electromagnetic and LFEM.
- Further experimental evaluation

## References

H. Chen and A.G. Cohn. Buried Utility Pipeline Mapping based on Street Survey and Ground Penetrating Radar, Proc. of the European Conference on Artificial Intelligence (ECAI'10), Lisbon, Portugal, 2010

H. Chen and A. G. Cohn. Probabilistic conic mixture model and its applications to mining spatial ground penetrating radar data. In Workshops of SIAM Conference on Data Mining (WSDM10), 2010.

H. Chen and A. G. Cohn. Probabilistic Robust Hyperbola Mixture Model for Interpreting Ground Penetrating Radar Data. In IEEE World Congress on Computational Intelligence. Barcelona, July, 2010.

www.mappingtheunderworld.ac.uk
www.comp.leeds.ac.uk/mtu

UNIVERSITY OF BIRMINGHAM · UNIVERSITY OF BATH · EPSRC Engineering and Physical Sciences Research Council · The University Of Sheffield. · University of Southampton · UKWIR