



# 3D mapping from partial observations: An application to utility mapping

Qingxu Dou<sup>a</sup>, Zhiyuan Lin<sup>b,\*</sup>, Derek R. Magee<sup>a</sup>, Anthony G. Cohn<sup>a,c,d</sup>

<sup>a</sup> School of Computing, University of Leeds, Leeds LS2 9JT, UK

<sup>b</sup> Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK

<sup>c</sup> Department of Computer Science and Technology, Tongji University, 1239 Siping Road, Shanghai, PR China

<sup>d</sup> School of Mechanical and Electrical Engineering, Qingdao University of Science and Technology, Qingdao, PR China

## ARTICLE INFO

### Keywords:

Utility mapping  
Partial observation  
Integer linear programming  
3D mapping

## ABSTRACT

Precise mapping of buried utilities is critical to managing massive urban underground infrastructure and preventing utility incidents. Most current research only focuses on generating such maps based on complete information of underground utilities. However, in real-world practice, it is rare that a full picture of buried utilities can be obtained for such mapping. Therefore, this paper explores the problem of generating maps from partial observations of a scene where the actual world is not fully observed. In particular, we focus on the problem of generating 2D/3D maps of buried utilities using a probabilistic based approach. This has the advantage that the method is generic and can be applied to various sources of utility detections, e.g. manhole observations, sensors, and existing records. In this paper, we illustrate our novel methods based on detections from manhole observations and sensor measurements.

This paper makes the following new contributions. It is the first time that partial observations have been used to generate utility maps using optimization based approaches. It is the first time that such a large variety of utilities' properties have been considered, such as location, directions, type and size. Another novel contribution is that different kinds of connections are included to reflect the complex layout and structure of buried utilities. Finally, for the first time to the best of our knowledge, we have integrated utility detection, probability calculation, model formulation and map generation into a single framework.

The proposed framework represents all detections using a common language of probability distributions and then formulates the mapping problem as an Integer Linear Programming (ILP) problem and the final map is generated based on the solution with the highest probability sum. The effectiveness of this system is evaluated on synthetic and real data using appropriate evaluation metrics.

## 1. Introduction

When managing massive urban underground infrastructure or to prevent incidents caused by underground working near utilities, it is required to have an accurate mapping of buried utilities such as gas pipes, water mains and telecommunication cables [1,2]. The statutory records obtained from the utility companies can provide such maps, but these records are notoriously inaccurate; for example it is reported [3] that in a survey of 187 incidents where utilities had been struck during excavation, 52% were not recorded in the utility company plans, and of the 89 that were on plans, 84% were inaccurately plotted. In the past ten years, utility incidents during excavation due to lack of record information caused 98 fatalities, 371 injuries and more than \$500 million total loss in the United States [4]. Moreover, the inaccuracy in the locations of utilities may also yield a large number of “dry holes” (i.e.

excavations where underground utilities are not found), resulting in waste of time, human resources and money. Instead of replying on statutory records, it is desirable to construct a utility map prior to commencing works based on on-site survey information such as sub-surface sensor measurements. There has been considerable research into mapping from sensor data (in particular work on SLAM [5,6]) in free space, where in principle further data may be collected to improve mapping. On the other hand, in some applications, it may be very expensive or technically too challenging to obtain sensor data for an entire scene, for instance, in mapping buried utilities. Hence developing mapping methods to cope with partial data are needed. We explore this problem in the context of mapping the underground network of utilities beneath the streets [7].

\* Corresponding author.

E-mail addresses: [Z.Lin@leeds.ac.uk](mailto:Z.Lin@leeds.ac.uk) (Z. Lin), [D.R.Magee@leeds.ac.uk](mailto:D.R.Magee@leeds.ac.uk) (D.R. Magee), [A.G.Cohn@leeds.ac.uk](mailto:A.G.Cohn@leeds.ac.uk) (A.G. Cohn).

<https://doi.org/10.1016/j.autcon.2020.103229>

Received 19 May 2019; Received in revised form 11 April 2020; Accepted 11 April 2020

Available online 23 May 2020

0926-5805/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### 1.1. Related work

The proposed research relates to several previous works on map generation, but none of them specifically addresses our research focus, i.e. (i) an exploration of the problem in a holistic probabilistic point of view to represent as many as possible real-world scenarios, (ii) a uniform framework for data integration and map generation, and (iii) complete algorithmic approaches for generating probabilities and for formulating an integer linear program to yield the final map. In addition, compared with some literature where only artificial data has been tested, we have both included real-world and synthetic data in our computational experiments. Where possible, we make comparisons between our research and the existing work while doing the survey.

First we consider previous work on SLAM [8–11]. A very basic difference between this work and SLAM is that in SLAM a map of an unknown environment is updated by keeping the track of an agent's location, so the sensor data should be recorded and used sequentially in time. In contrast, in this work, the sensor data can be captured in any order and the whole map is generated at a single time by combining all the observations.<sup>1</sup> In addition, SLAM focuses on improving the accuracy of the measurements of sensor data and then obtaining a better map of the unknown environment while this work focuses on inferring the connections between utility detections, and their location.

Second, this work is related to existing work in the area of buried utility mapping [12–14]. [12,13] focus on sensor data fusion and do not use any information from street furniture surveys in the mapping. The system proposed by [14] is for creating a map of buried utilities based on manhole observations. However in [14], only straight pipes between manholes are taken into account while in the system proposed here, direct-connections and side-connections are included in the created map. To the best of our knowledge, this is the first time that utility properties other than location and direction are taken into account when generating maps of buried utilities.

An information fusion approach that integrates both sensing and non-sensing data based on Dempster-Shafer (D-S) evidence theory is presented in [2]. Their experiments suggest that the integration of sensing and non-sensing data significantly reduces the error of estimated location and enhance the confidence of the results. There is potential to also incorporate non-sensing data [1] such as existing records, utility specifications in our proposed framework, as long as these data can be represented by appropriate probability distributions.

Our research also relates to other works on underground utility mapping [15–18]. In [15], the authors present the initial results of a framework for 3D mapping of underground utilities especially an overview of methods for primary data capture. A framework is proposed in [16] for utility data governance from the underground utility data survey to data usage, while [17] proposes an approach that integrates multiple available utility location datasets to represent geographical uncertainties. In [18], registration of utility networks in a cadastre is considered in a 3D + time (=4D) context. Unlike the aforementioned works, which were mainly focused on data capture and 3D data registration, our work explores the holistic theoretical problem, gives a comprehensive framework for both data integration and map generation, as well as proposing a complete algorithmic procedure for generating the probabilities and formulating the integer programming model. All this shows the originality of and need for our research from both a practical and a theoretical point of view.

### 1.2. Problem statement and methodology

This work presents a framework to construct a most probable map of buried utilities based on data from street furniture surveys and

<sup>1</sup> This does not exclude the possibility of taking further sensor measurements and rerunning the framework of course

subsurface surveys with multiple sensors.

In street furniture surveys, the information of manhole locations and the depth and direction of each pipe or cable end going through manholes is collected. In subsurface surveys, multiple sensors such as GPR [7,19–21] or a Vibro-acoustic sensor (VA) [22] are used to capture sensor measurements. For a utility end observed from a manhole, based on an assumption that utilities follow a linear route, the direction of this utility is described by a probability field. If a hypothesized detection (HD) extracted from sensor measurements is a response of a utility, it also indicates the going direction of this utility. The probability field related to this utility end is updated by HDs associated with it. In this work, we employ *Integer Linear Programming* (ILP) to solve the mapping problem. ILP can take associations of any utility end with any subset of the HDs. This use of ILP guarantees that the most probable map is generated. Algorithms are proposed for computing the probability of connecting any two utility ends by extending them (referred to as direct-connections) and the probability of extending a third utility end so as to connect it to a utility segment formed by direct-connection (referred to as side-connection). Fig. 1 gives two examples of how direct and side connections can be formed. Besides the locations and directions of utilities, other properties of them such as type and size are also used in the probability computation. Hypothesized detection free measurements (HDFs) are also incorporated into this framework to modify the probability of connecting two utility directions.

### 1.3. Structure of the paper

The rest of the paper is organized as follows: The framework of the proposed method to generate a 3D map of buried utilities is discussed in Section 2, which is followed by the probability computations for utility connections in Section 3. The method of generating 3D maps by solving the formulated ILP is described in Section 4. Experimental results are shown and analysed in Section 5 and the paper finishes with conclusions in Section 6.

## 2. Overview of proposed framework

In this section, we give an overview of the proposed framework.

### 2.1. Manhole observations

From street furniture surveys, the locations of manholes are recorded and the manhole covers are lifted to check utilities going through them. For an observed utility end from a manhole, its depth is measured and the direction the utility takes as it leaves the manhole is estimated. Multiple utility ends may be observed from a single manhole so a manhole with utility ends observed in it can make up multiple manhole-utility pairs that are used for further processing.

In this work, we assume that the manhole locations and the start points of the observed utility ends are measured accurately (e.g. using a total station theodolite). Given a manhole-utility pair, under the linearity assumption, a utility goes out from the start point along a certain direction in a straight line. Since the directions of the utility ends are observed manually, they are not accurate. The going direction of a utility is described by a probability distribution (assumed to be Gaussian in this work).

### 2.2. Sensor measurements

Sensor measurements captured in subsurface surveys [7,23] can be incorporated into the proposed framework. From a sensor measurement hypothesized detections (HDs), which reflect the presence of certain utilities can be extracted [24–29] from the raw sensor data.

Hypothesized detection free measurements (HDFs) are also incorporated in the proposed framework. An HDF is a cross section (see detailed definition in Section 3) where a measurement was performed

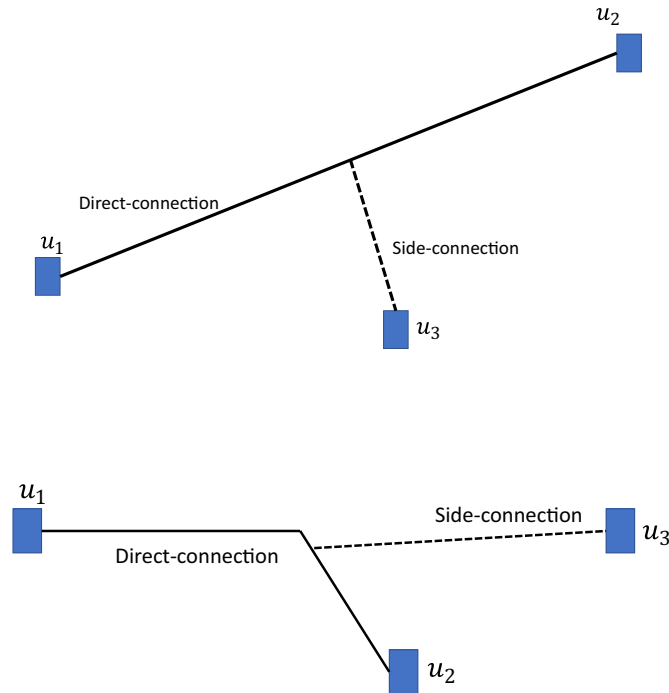


Fig. 1. Two exemplary cases (top and bottom) of direct-connection (solid line joining  $u_1$  and  $u_2$ ) and side-connection (dashed line joining  $u_3$  to the existing pipe segment formed by direct-connection between  $u_1$  and  $u_2$ ).

but no HD was extracted. Note that even there is no HD, that does not mean there is no utility, since sensors do not always detect buried utilities, for example because a GPR signal is attenuated severely in clay.

### 2.3. Connection probability calculation and integer linear programming model

After the data from manhole inspections has been collected and the HDs from sensor measurements have been extracted, utility ends, HDs and HDFs are associated to make up *udf-combinations*. The term “udf” comes from the three involved entities in such a combination: a utility end, hypothesized detections and hypothesized detection free measurements.

Based on the proposed algorithms (see details in Section 3), the probabilities of connecting the utility ends with direct-connections and side-connections are computed.

Next, we feed the computed probabilities into an Integer Linear Programming model given in Section 4. By maximizing the sum of connecting probabilities, a map of the buried utilities is obtained. The advantage of using integer linear programming is that all theoretically possible scenarios can be considered in a single model subject to certain constraints that make sure the resulting optimal mapping also complies with real-world requirements.

### 2.4. A flow chart of the framework

Fig. 2 gives a flow chart of the proposed framework. First, the input data is captured from manhole observations and sensor measurements. Second, HDs and HDFs are associated to observed utility ends to make up *udf-combinations*. After the probabilities of any direct-connections and side-connections are computed, an integer linear programming (ILP) model is employed to find the most probable connections.

## 3. Probability computation for utility connections

The direction uncertainty of a buried utility can be reflected by a probability distribution of its going direction. The Gaussian distribution is chosen in this work. One advantage of using Gaussian distributions is that they can be conveniently combined to form a compound distribution by taking account of several data sources, e.g. manhole observations and sensor measurements as in our work. Other types of sources, such as existing maps, CAD drawings and utility specifications [2] can potentially be formulated into the proposed framework, as long as the data sources can be described by an appropriate probability model. In each of the subsection below this section, we present how to compute the probabilities of direct-connections and side-connections in different situations. Alongside the technical exposition of the computation of these probabilities, we add illustrative figures to help better explain how these computations can be applied in practice.

### 3.1. The probability distribution of a utility end

Based on the street furniture survey, the properties recorded for each utility  $u_n$ ,  $1 \leq n \leq N$  includes the coordinates of the start point  $S_n$  of the utility end, the direction of the observed utility end represented by two angles **pan** and **tilt**<sup>2</sup> as represented by  $\alpha_n$  and  $\beta_n$  in Fig. 3 and other properties such as its size and type. Since a typical total station theodolite used for geo-measurement can measure distances with an accuracy of about 1.5 mm over a distance of up to 1500 m [30], it is assumed that the start points of the observed utility ends are measured accurately. On the other hand, the directions of the utility ends, which are observed manually, are usually not recorded accurately. As can be seen in Fig. 3, an arrow with a dotted line represents an observed utility direction and other arrows starting from the same point represents other possible going directions of this utility. In this work, it is assumed that the utility direction obeys a Gaussian distribution with respect to the random vector (**pan**, **tilt**) with the observed direction as its most likely extending direction. For a utility end with observed direction  $(\alpha_n, \beta_n)$ , the probability of it going along the direction represented by  $(\alpha, \beta)$  is estimated by the following probability density:

$$p_{ld_n}(\alpha, \beta, \alpha_n, \beta_n) = \frac{1}{2\pi |\Sigma_n|} \exp\{-(A_u - A_n)^T \Sigma_n^{-1} (A_u - A_n)\}, \quad (1)$$

where  $A_u = (\alpha, \beta)$ ,  $A_n = (\alpha_n, \beta_n)$  with  $T$  representing vector transpose.  $\Sigma_n \in \mathbb{R}^{2 \times 2}$  represents the covariance matrix of the Gaussian distribution of the  $n$ -th observed utility end. The covariance matrix of each observed utility end is recorded as a property of it, which reflects how accurately the utility direction is estimated. In practice, the parameters in the covariance matrix can be learned in an area with ground truth and then applied in other areas with similar conditions. Note that the accuracy may depend on the person performing the estimation, and a learned model may thus not be as useful if the person using the model has a different estimation model. This is unlikely to be a problem in practice though.

### 3.2. Associating utility ends with HDs and HDFs

Besides the street furniture surveys, the hypothesized detections (HDs)  $H = \{h_m\}$ ,  $1 \leq m \leq M$  extracted from sensor measurements and hypothesized-detection-free measurements (HDFs)  $F = \{f_k\}$ ,  $1 \leq k \leq K$  are also very important in the utility map generation. For each hypothesized detection  $h_m$ , based on the geo-measurement of the related sensor measurement and the intrinsic properties of the related sensors, the location of  $h_m$  is computed. Given a threshold value  $\theta_p$  and the

<sup>2</sup> **pan** is the horizontal angle of a vector counter-clockwise from the positive direction of the x-axis and **tilt** is the vertical angle between a vector and the positive direction of the z-axis.

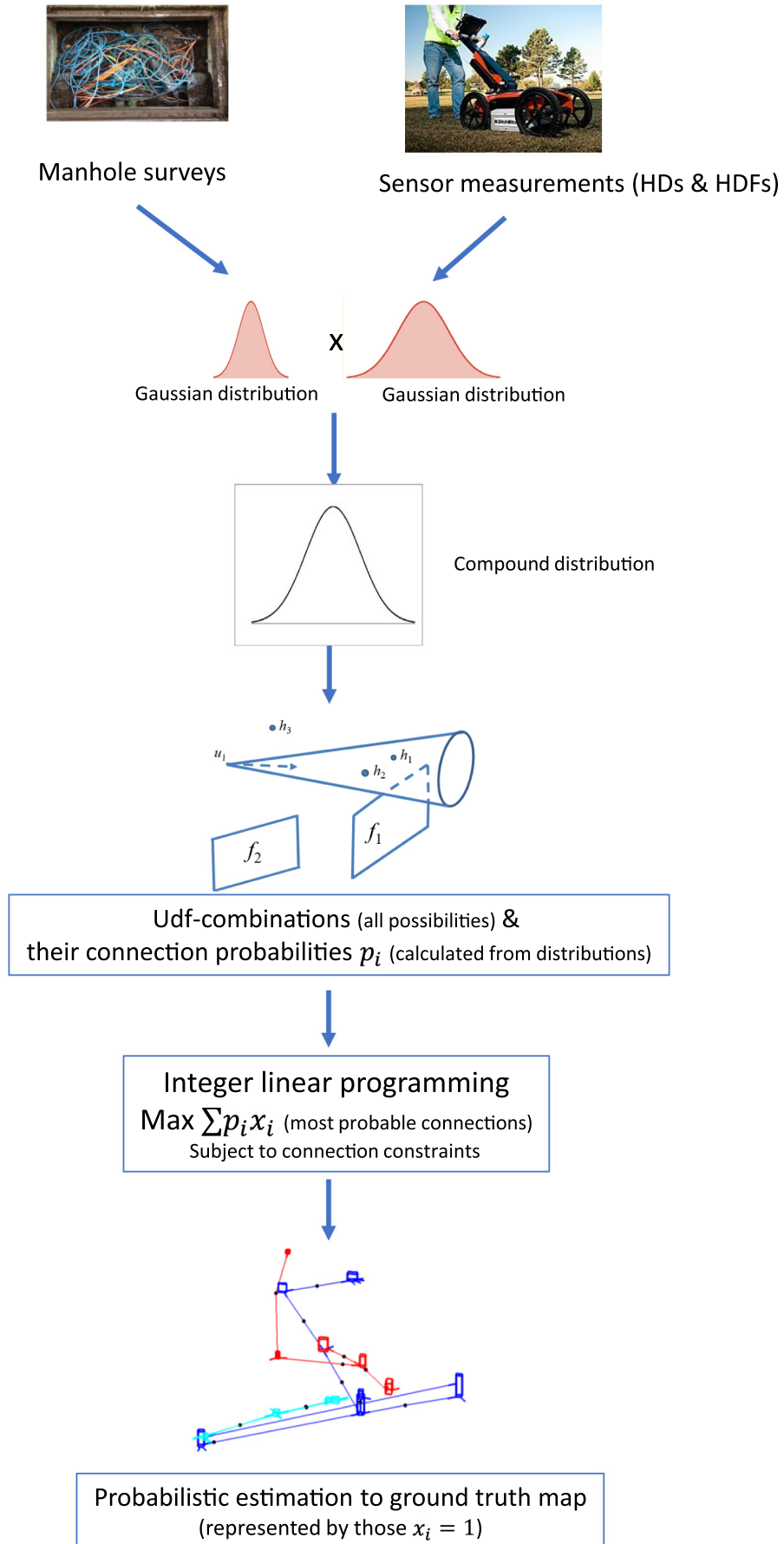


Fig. 2. Flow chart illustrating the framework proposed in this paper (source of the image of a GPR unit: [https://en.wikipedia.org/wiki/Ground-penetrating\\_radar](https://en.wikipedia.org/wiki/Ground-penetrating_radar)).

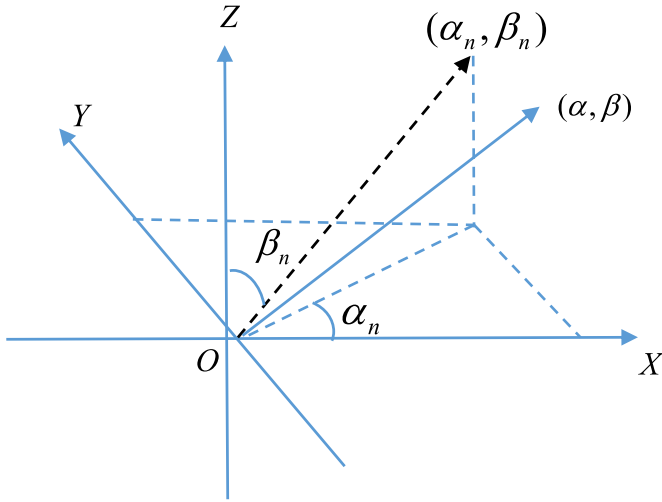


Fig. 3. The direction of a utility is represented by two angles pan and tilt denoted as the angles  $\alpha_n$  and  $\beta_n$  in this figure. The observed utility direction is represented by an arrow with a dotted line and other possible going directions of the utility are represented by arrows with a solid line.

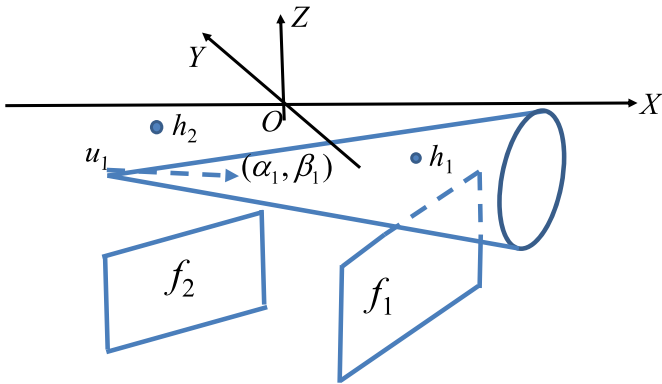


Fig. 4. This figure illustrates how to associate a utility end ( $u_1$ ) with HDs and HDFs. Given a threshold value to the probability field generated by an observed utility end with respect to the going direction, a cone with the start point at the apex, and observed direction as the axis is formed. HDs inside of the cone such as  $h_1$  are associated with the utility cone. If the cross section defined by an HDF has an intersection with the cone such as  $f_1$ , it is also associated with the utility end.

probability field defined by Eq. (1), the probability field with probability values greater than  $\theta_p$  forms a cone with the start point of the utility as the apex and the observed direction of this utility end as the axis (see Fig. 4). The HDs inside of the cone generated by a specific utility end are associated with the related utility, i.e. they are regarded as potential responses from this utility. Theoretically, ILP can deal with the associations of any utility ends with any HDs. However, if a HD is too far from a utility, it does not make any sense to associate them. To reduce the computation load, we use a utility cone to limit the number of associations between utility ends and HDs. After all  $h_m \in H$  are associated with different observed utility ends, a utility end may have several  $h_m$  associated with it and at the same time an  $h_m$  may be associated with multiple utility ends.

Among the provided sensor measurements, besides the ones with some hypothesized detections extracted from them, there are usually some measurements without any hypothesized detections extracted. Such measurements are hypothesized-detection-free measurements as aforementioned. In this work, HDFs are also included in the proposed system. A HDF is represented by a cross section, which is a rectangle with one side being the scan line of the sensor and two sides

perpendicular to the ground surface with the length equal to the depth the sensor can measure. Two examples of HDFs ( $f_1$  and  $f_2$ ) are presented in Fig. 4. With respect to the same probability threshold  $\theta_p$ , if the cross section of an HDF has intersection with a probability cone, the HDF is associated with the related utility end (e.g.  $f_1$  in Fig. 4). A utility end may have multiple HDFs associated with it and an HDF may be associated with multiple utility ends.

If a utility end  $u_n$  has multiple HDs associated with it, due to the complexity of the utility configuration, the sensor measurement error and the processing error in the extraction step, an HD associated with a utility may not be a response of this utility. In this work, among all the HDs associated with a utility, we test all possible combinations. For example, if  $u_1$  has two HDs  $h_1$  and  $h_2$  associated with it, there are four different combinations, i.e. none of them is a response of  $u_1$ , only one of them is a response of  $u_1$  and both of them are responses of  $u_1$ . A utility with any subset of the HDs associated with it and all the HDFs associated with it make up a combination called a *udf-combination*.

### 3.3. Using extracted hypothesized detections

For a specific *udf-combination*, the utility end  $u_n$  in this *udf-combination* generates a probability distribution, which is a Gaussian and can be computed with Eq. (1).

Suppose there are HDs in this *udf-combination*, then for a specific HD  $h_m$  in this *udf-combination*, the vector connecting the start point of  $u_n$  and  $h_m$  indicates the direction of the utility under the linear assumption. Representing the vector connecting the start point of  $u_n$  and  $h_m$  as  $v_{mn}$ , similar to the probability distribution of the direction of a utility generated by an observed utility end, the start point of  $u_n$  with vector  $v_{mn}$  generates a probability distribution with respect to the directions of the related utility end. In this work, a Gaussian is also used to describe this probability distribution. The covariance matrix set for the Gaussian distribution reflects the accuracy of the related sensor. Thus, we have two Gaussians associated with the utility end  $u_n$ . These two distributions can be merged into an improved estimate by multiplying them together [31]. As pointed out by [31], merging two Gaussian distributions in this way leads to reduced standard deviations. Since multiplying Gaussian distribution results in a Gaussian distribution, the operation is symmetric, associative, and can combine any number of distributions in any order. So for a *udf-combination* with multiple HDs, the final distribution of this *udf-combination* is the product of all the probability distributions generated by the utility end and each HD in this *udf-combination*. With respect to the final distribution, the ray, along which the probability has the largest value, is called the axis of the distribution of a *udf-combination*.

### 3.4. Probability of connecting two Udf-combinations based on location and direction

For two *udf-combinations* generated from the same utility end, the probability of connecting these two *udf-combinations* is obviously 0. For two *udf-combinations* generated from different utility ends, based on the assumptions that the probability distributions of two *udf-combinations* generated from different utility ends are independent to each other and the utilities are linear, the probability of connecting two utility ends by direct-connection based on location and direction at a certain point  $I$  is the product of the probabilities of these two utilities when both of them go through point  $I$ . The probability of connecting two utility ends with a direct-connection based on location and direction can be defined as follows:

$$pld_{i,j} = \max_I (pld_i \cdot pld_j) \quad (2)$$

This means that to compute the probability of connecting two utility ends  $u_i$  and  $u_j$  by direct-connection based on location and direction we should find the connecting point  $I_{i,j}$  where the product of  $pld_i$  and  $pld_j$



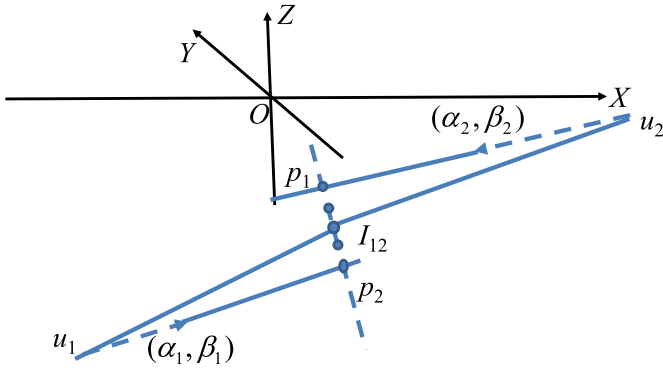


Fig. 5. A diagram illustrating how to compute the probability of connecting the utility ends ( $u_1, u_2$ ) in two udf-combinations by direct-connection with respect to a very common situation where the lines passing through the start points of the utility ends and going along the axes of the probability distributions of the udf-combinations are skew lines and there exists a common perpendicular line to these two skew lines intersecting with the extensions of the utility ends along the axes of the udf-combinations.

obtains its maximum value. In the following section, methods to estimate the location of  $I$  in different scenarios are presented with  $u_1$  and  $u_2$  as the examples.

A scenario often encountered is that the lines passing through the start points of  $u_1$  and  $u_2$  along the axes of the related udf-combinations are skew lines and at the same time there is a common perpendicular of these two skew lines intersecting with the extensions of  $u_1$  and  $u_2$  along the axes at  $P_1$  and  $P_2$  respectively as represented by Fig. 5. In this situation, the common perpendicular line segment  $P_1P_2$  is discretized first with respect to the resolution of angles used for describing the direction of the utilities. For each of the discretized points, the probabilities of  $pld_1$  and  $pld_2$  are computed with Eq. (1) and then their product can be obtained. The maximum product is recorded as  $pld_{1,2}$ , the probability of connecting two utility ends by direct-connection based on their location and direction, and the related point is recorded as  $I_{1,2}$ .

Besides the scenario described above, there are several special scenarios as represented in Fig. 6. The first one is that two utilities intersect each other at a point  $I_{1,2}$  after the utility ends are extended along the corresponding axes of udf-combinations as represented with Fig. 6(a). In this scenario the probabilities of  $pld_1$  and  $pld_2$  have the largest value with respect to the corresponding probability distributions so their product has the largest value and the point  $I_{1,2}$  is recorded as the intersection point of these two utilities. In the scenarios represented in Fig. 6(b) and (c), the extension of one of the two utility ends along the axis of the related udf-combination goes through the start point of the other utility end or intersects at a point on the reverse extension of the other utility end. In both scenarios, the probabilities of connecting these two utility ends by direct-connection based on location and direction is computed with respect to the line segment between the two start points and the mid-point of this line segment is recorded as the intersection point. Fig. 6(d) describes a scenario where the lines passing through the start points and going along the observed directions are skew lines with one of their common perpendicular intersecting with one on the extension of the utility end along the axis of the udf-combination and the other one on the reverse extension of the utility end with respect to the udf-combination axis. A very special case of the scenario represented by Fig. 6(d) is that the common perpendicular line intersects with one of the utilities at its start point. In this scenario, the probability of connecting these two utility ends is also computed with respect to the line segment between the two start points.

### 3.5. Using hypothesized-detection-free (HDF) measurements

Just as a hypothesized detection indicates the presence of a utility to

a certain extent, HDF measurements correspondingly indicate the absence of utilities to some extent. If a utility is supposed to pass through the cross section CS of an HDF with certain probability, since no hypothesized detection is extracted from this HDF, the probability of the related utility should be reduced. The attenuation ratio reflects our confidence on the related sensor. In this situation, the smaller the attenuation ratio is, the more confidence we have in the related sensor.

For each related connection, we try to find the maximum probability value of it to use in the Integer Linear Programming formulation. For a connection with the largest probability, if the connected utility segment passing through a CS of an HDF, the probability is attenuated and then the related connection may not be the one with the maximum probability value any more. The reason is that, besides the connections passing through the related CS, there are some connections not passing through this CS and the related probability values are not attenuated. In a probability distribution generated by a udf-combination, the maximum probability is obtained along the axis of this udf-combination; for the probability values along other rays, the closer the ray is to the axis, the larger the probability value that can be obtained. So for the connections not passing through the related CS, the maximum probability value can only be obtained when the two utilities are connected at a point on the boundary of the CS. So in this work, if a connection passes through a CS of an HDF measurement, the attenuated maximum probability is compared with the probabilities when the utility ends are connected at any point on the boundary of the CS. If there are some probabilities larger than the attenuated probability, the largest one is recorded as the probability of connecting these two utility ends and the related point on the boundary of CS is recorded as the connection point as the point  $B$  shown in the Fig. 7.

### 3.6. Using utility properties other than location and direction

From the street furniture survey, besides the start point location and direction of an observed utility, other properties such as its size and type may be also recorded. The recorded properties other than location and direction of the observed utilities are also formulated in the proposed system.

It can be seen that the probability of connecting two udf-combinations by direct-connection given by Eq. (2) is only related to the locations and directions of the axes of the corresponding udf-combinations. Obviously, when connecting two utility ends, the other properties such as their size and type should be taken into account: for example if two utility ends have different types (e.g. one is *Water main* and the other one is *Cable*), there is no possibility to connect them. Thus we define the probability of connecting utility ends  $u_i$  and  $u_j$  based on their types as follows:

$$pt_{i,j} = \begin{cases} 1, & \text{if types are same} \\ 0, & \text{if types are different} \\ 1/k, & \text{if one or both of the types is unknown, where } k \\ & \text{is the number of unknown types} \end{cases} \quad (3)$$

Besides the type of a utility, the size of a utility i.e. the diameters of the pipes or cables is another important property. Considering that the Laplace distribution has a sharp peak and heavier tails than a Gaussian distribution, a Laplace distribution is selected to compute the probability of connecting two utilities based on the size of them. First, the sharp peak of Laplace distribution can benefit the connection of two utilities with similar size. Second, the heavier tails of Laplace distribution can tolerate large measurement errors. For two utility ends  $u_i$  and  $u_j$ , if the recorded diameters of them are  $r_i$  and  $r_j$  respectively, the probability  $ps_{i,j}$  can be estimated as follows:

$$ps_{i,j}(r_i, r_j, b) = \frac{1}{2b} \exp\left(-\frac{|r_i - r_j|}{b}\right) \quad (4)$$

where  $b > 0$  is a scalar parameter of a Laplace distribution. The

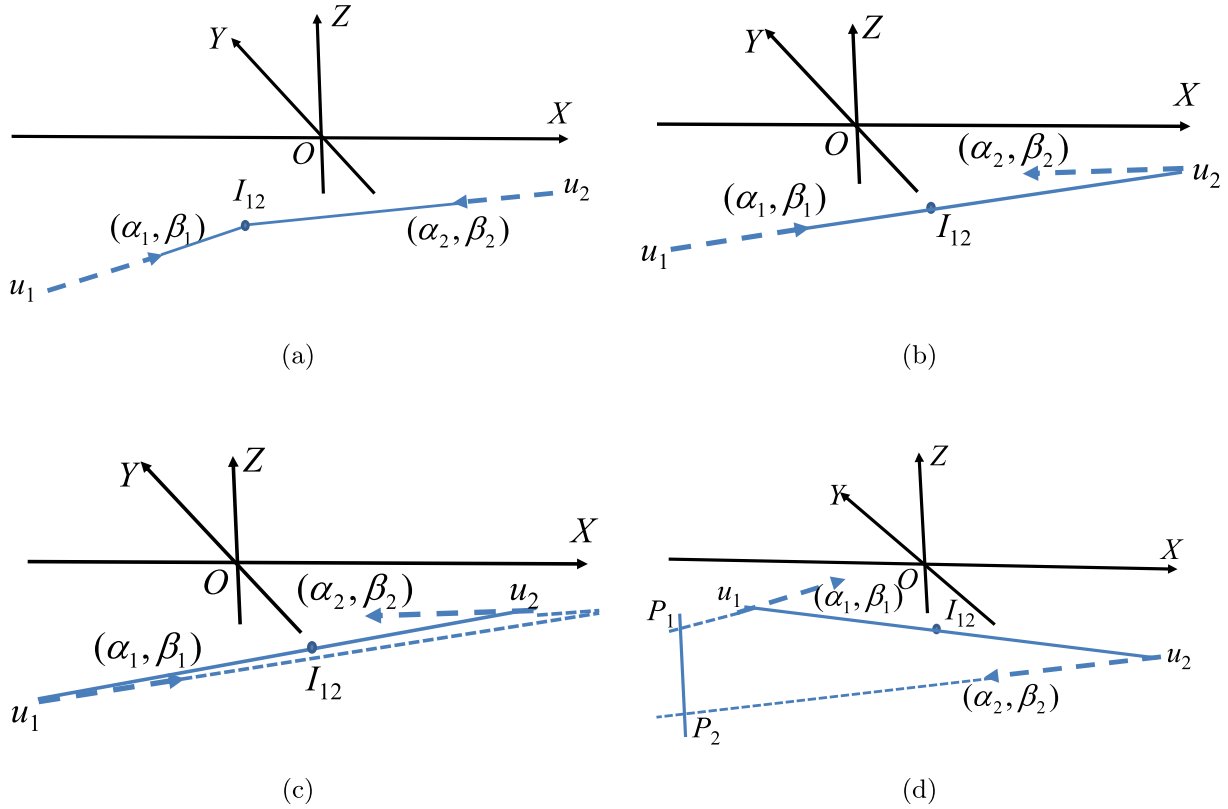


Fig. 6. Different scenarios when computing the probabilities of connecting two observed utility ends ( $u_1, u_2$ ) with direct-connections based on location and direction.

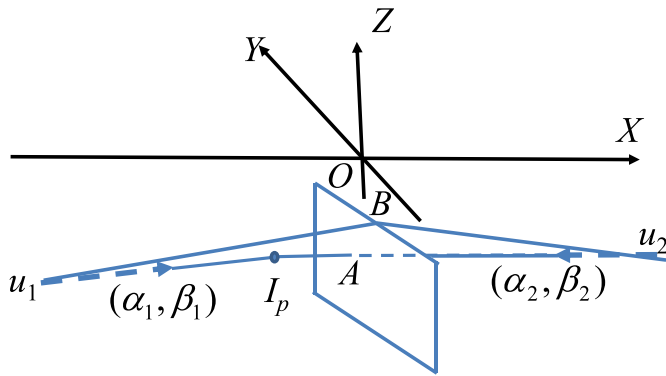


Fig. 7. A diagram illustrating how to use an HDF measurement. If a connected utility segment passes through the cross section of an HDF measurement, then the related probability is attenuated and the attenuated probability is compared with the probabilities when connecting these two utility ends at any point on the boundary of this cross section; then the connecting point  $I_p$  is chosen according to the maximum probability value (see detailed explanation in text).

smaller the value of  $b$ , the larger is the value of the peak of the Laplace distribution.

If other properties such as materials of the utilities are recorded, the associated probabilities can be defined and taken into account. In the experiments reported below in Section 5, besides the direction and location of utilities, only the type and size are taken into account, therefore only the probabilities of these properties are defined explicitly in this paper.

### 3.7. Probability of connecting two utility ends with direct-connection

After the probabilities of connecting the utility ends of two udf-combinations based on location and direction and other properties are

computed, the probability of connecting two utility ends with direct-connection  $pd_{i,j}$  can be computed. The type of a utility and the location and direction of this utility are independent to each other and within a certain type of utility, the size of a utility is independent of its location and direction, so the probability of  $pd_{i,j}$  can be defined as the product of  $p_{ld_{i,j}}$ ,  $p_{t_{i,j}}$  and  $p_{s_{i,j}}$ , i.e.

$$p_l = pd_{i,j} = p_{ld_{i,j}} \cdot p_{t_{i,j}} \cdot p_{s_{i,j}} \quad (5)$$

### 3.8. Probability of connecting a utility end to a utility segment by side-connection

After the utility ends of two udf-combinations are connected by direct-connection to make a utility segment between two manholes, the connection of the utility end of a third udf-combination to this utility segment is called a side-connection as represented in Fig. 8.

Let  $l$  denote the direct-connection of two utility ends and let  $c$  represent a side-connection of a third utility end to the utility segment made by the direct-connection. The probability of connecting a utility end  $u_k$  to a utility segment formed by connecting  $u_i$  and  $u_j$  with direct-connection can be represented as follows

$$p_c^l = ps_{k \rightarrow i,j} = p(E_{k \rightarrow i,j} | E_{i,j}) \cdot p(E_{i,j}) \quad (6)$$

where  $E_{i,j}$  represents the event of connecting  $u_i$  and  $u_j$  with direct-connection and  $E_{k \rightarrow i,j}$  represents the event of connecting  $u_k$  to the segment formed by  $u_i$  and  $u_j$ . The probability  $p(E_{i,j})$  is given by Eq. (5). When computing the probability of  $p(E_{k \rightarrow i,j} | E_{i,j})$ , since  $u_i$  and  $u_j$  have been connected before connecting  $u_k$  to this segment, the location of this segment is fixed. Suppose the utility segment formed by  $u_i$  and  $u_j$  is connected at point  $I$  as depicted by Fig. 8, and then if utility  $u_k$  is connected to the segment formed by  $u_i$  and  $u_j$  by side-connection, it connects to  $S_i I$  or  $S_j I$  at some point  $T$ . To decide the location of  $T$ , the common perpendicular of the ray starting from  $u_k$  along the axis of the corresponding udf-combination and  $S_i I$  or  $S_j I$  is computed. If the

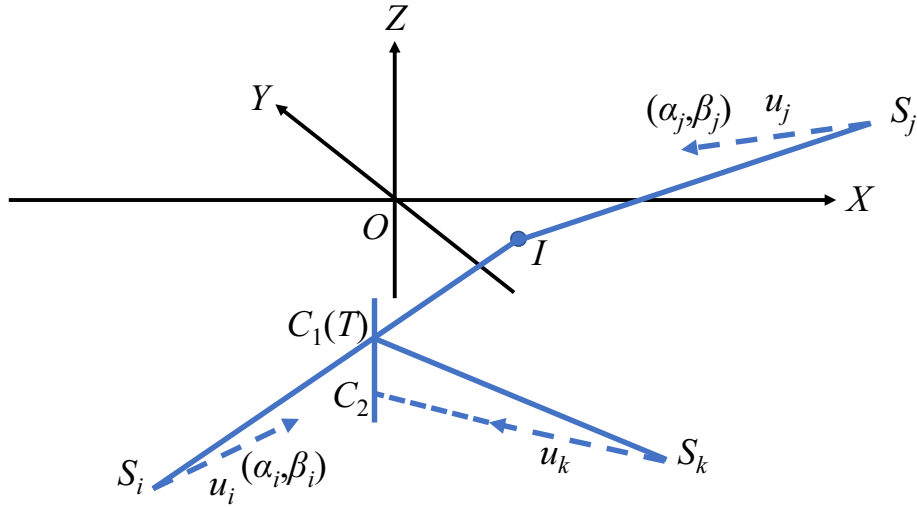


Fig. 8. A diagram illustrating connecting a utility end  $u_k$  by side-connection to a utility segment obtained by direct-connection of  $u_i$  and  $u_j$ . In this figure,  $S_i$ ,  $S_j$  and  $S_k$  are the start points of the corresponding utilities.

common perpendicular, which intersects with the ray and  $S_i$  or  $S_j$ , exists as  $C_1C_2$  depicted in Fig. 8, then the intersection point of this common perpendicular with  $S_i$  or  $S_j$  is the connection point  $T$ . The probability  $p(E_{k \rightarrow i, j} | E_{i, j})$  can be computed as

$$p(E_{k \rightarrow i, j} | E_{i, j}) = pl_{d_{k \rightarrow i, j}} \cdot pt_{k \rightarrow i, j} \cdot ps_{k \rightarrow i, j} \quad (7)$$

where  $pl_{d_{k \rightarrow i, j}}$  represents the probability of connecting  $u_k$  to the segment formed by  $u_i$  and  $u_j$  just based on the location and direction and  $pt_{k \rightarrow i, j}$  and  $ps_{k \rightarrow i, j}$  are the probabilities computed based on the type and size properties of  $u_k$  and  $u_i$  or  $u_j$ . If the common perpendicular line segment, which intersects with the ray generated by  $u_k$  and the segment formed by  $u_i$  and  $u_j$ , does not exist, and then the probability  $pl_{d_{k \rightarrow i, j}}$  is set to 0.

#### 4. 3D map generation by integer linear programming

Integer Linear Programming (ILP) [32] is a mathematical optimization technique that in the general case can be described using the following canonical form with  $n$  variables and  $m$  constraints:

$$\text{minimize} \quad \sum_{j=1}^n c_j x_j \quad (8)$$

$$\text{subject to} \quad \sum_{j=1}^n a_{ij} x_j \leq b_i, \quad i = 1, \dots, m \quad (9)$$

$$x_j \in \mathbb{Z}_+. \quad (10)$$

In the above formulation,  $\sum_{j=1}^n c_j x_j$  is called the objective function and  $\sum_{j=1}^n a_{ij} x_j \leq b_i$ ,  $i = 1, \dots, m$  are the constraints, both of which should be linear. The decision variables  $x_j$ ,  $j = 1, \dots, n$  are required to be non-negative integers. A special case of ILP is when the decision variables are binary, i.e.  $x_j \in \{0, 1\}$  as is the case in this work.

The goal of ILP is to minimize (as in the canonical form) the objective function while satisfying all constraints and variable domain requirements. ILP is known to be NP-hard [33]. However, in our research, the problem sizes are rather small: after removing variable redundancy, our synthetic instance has 3538 binary variables and 3235 constraints, and the real-world instance has 50 binary variables and 22 constraints. Therefore it is sufficient to use a state-of-art commercial ILP solver to solve the problem by branch-and-bound.

In a solution of the ILP of this work, for a probability value  $c_j$  of connecting two utility ends by direct-connection, only if the related decision variable  $x_j$  is 1 are the two utility ends then connected by direct-connection. For a probability value  $c_j$  of connecting a utility end to

a utility segment with side-connection, only if the related decision variable  $x_j$  is 1 is this utility end connected to the utility segment. The final utility map is generated by connecting the utility ends indicated by all the decision variables with value 1 with the corresponding connection option, direct-connection or side-connection.

To decide which combinations should be used in the connections and how they are connected to each other, an Integer Linear Programming formulation is designed for solving the aforementioned mapping problem.

For any two combinations  $c_i, c_j \in C$ , the probability of connecting them by direct-connection to make a utility segment between two manholes is computed. A special case is that a udf-combination may not be connected to any other combinations, but is extended to the boundary of the working area. For each such udf-combination, it is assumed that the udf-combination is connected to an artificial boundary udf-combination, which is denoted as  $c_b$ . A utility segment formed by  $c_i$  and  $c_j$  by direct-connection is denoted as  $l(c_i, c_j)$  and the set of all utility segments formed by direct-connection of two combinations is denoted as  $L = \{l(c_i, c_j) | c_i \in C, c_j \in C \cup \{c_b\}\}$ . Let  $p_l$  denote the connection probability of a direct-connection  $l(c_i, c_j) \in L$  and  $C(u_n)$  denote the set of the combinations generated from utility end  $u_n$ . Since two combinations from the same utility end cannot make a utility segment, we set  $p_{l(c_i, c_j)} = 0$  if  $c_i, c_j \in C(u_n)$  for any  $u_n \in U$ . If one of  $c_i$  and  $c_j$  is  $c_b$ , then the related  $p_l$  is set to a small constant value to ensure the correctness of the normal cases is not affected. For other situations, the computation of  $p_l$  is given in Sections 3.4, 3.6 and 3.7.

After the probabilities of  $p_l$  are computed for any  $l \in L$ , the probabilities of  $p_c^l$  of connecting a udf-combination  $c \in C$  to a direct connection  $l \in L$  is computed. Details of the computation are given in Section 3.8.

As for the decision variables, let  $x_l \in \{0, 1\}$ ,  $\forall l \in L$  be the binary variables for the direct-connections such that  $x_l = 1$  if direct-connection  $l$  is used in the solution and  $x_l = 0$  otherwise. Binary decision variables  $y_c^l \in \{0, 1\}$ ,  $\forall c \in C, \forall l \in L$  are used to determine the side-connections such that  $y_c^l = 1$  if a side-connection from  $c$  to direct-connection  $l$  is used in the solution, and  $y_c^l = 0$  otherwise.

We now formally present the ILP formulation for the aforementioned utility mapping problem, thus:

$$\text{maximize} \quad \sum_{l \in L} p_l x_l + W \sum_{c \in C, l \in L} p_c^l y_c^l \quad (11)$$

$$\text{subject to:} \quad y_c^l \leq x_l, \quad \forall l \in L, \forall c \in C \quad (12)$$



$$\sum_{c \in C(u_n)} \left( \sum_{q \in C \cup \{c_b\}} x_{l(c,q)} + \sum_{l \in L} y_c^l \right) \leq 1, \quad \forall u_n \in U \quad (13)$$

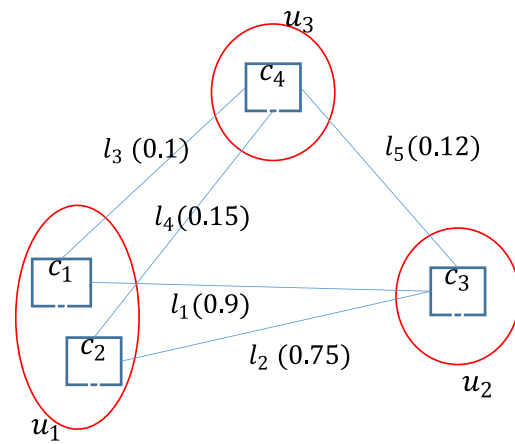
$$x_l \in \{0, 1\}, \quad \forall l \in L \quad (14)$$

$$y_c^l \in \{0, 1\}, \quad \forall c \in C, \forall l \in L \quad (15)$$

The objective function (11) maximizes the total probability given by the connection, including the ones contributed from direct-connections as well as from side-connections.  $W$  is a weight to penalize the influence from side-connections as usually they are less important than direct-connections. Constraints (12) make sure that a side-connection to a direct-connection can only be triggered if the direct-connection is used in the first place. Constraints (13) state that for each utility end at most one of its associated udf-combinations can be used in the solution, and when a udf-combination is used, it can only choose one connection type either as a direct-connection (i.e.  $\sum_{c_l \in C \cup \{c_b\}} x_{l(c,c_l)} = 1$ ) or as a side-connection (i.e.  $\sum_{l \in L} y_c^l = 1$ ). Finally, (14) and (15) give the domain of the decision variables.

To generate a map of the selected area, the observed manholes and the utility ends observed from each manhole are drawn. Once the solution of Objective (11) is obtained by solving the Integer Linear Programming problem, the direct-connections indicated by parameter  $x_l$  are connected at the corresponding connection point  $l$ . Then the parameter  $y_c^l$  is used to pick out the udf-combinations, which are connected to certain  $l$  by side-connections at the related connection point  $T$ . The final map includes the direct-connections and side-connections.

We illustrate how the above ILP model is formulated by a toy example given in Fig. 9. Three utility ends  $u_1, u_2$  and  $u_3$  associated with four udf-combinations  $c_1, \dots, c_5$  are given, where  $u_1$  is linked with  $c_1$  and  $c_2$ ,  $u_2$  is linked with  $c_3$  and  $u_3$  is linked with  $c_4$ . The five direct-connections are marked by  $l_1, \dots, l_5$  with their corresponding probabilities in the parentheses afterwards. The six side-connections are listed in the table with their probabilities. For simplicity, the boundary combination  $c_b$  is ignored here, which can always be added back if needed. Based on this toy example, the following ILP model can be formulated:



- utility ends
- udf-combinations

$l_i(p_{l_i})$  direct-connections with probabilities

$c$	$l$	$p_c^l$
$c_1$	$l_5$	0.01
$c_2$	$l_5$	0.02
$c_3$	$l_3$	0.1
$c_3$	$l_4$	0.2
$c_4$	$l_1$	0.9
$c_4$	$l_2$	0.5

Fig. 9. A toy example of how to formulate the ILP model.

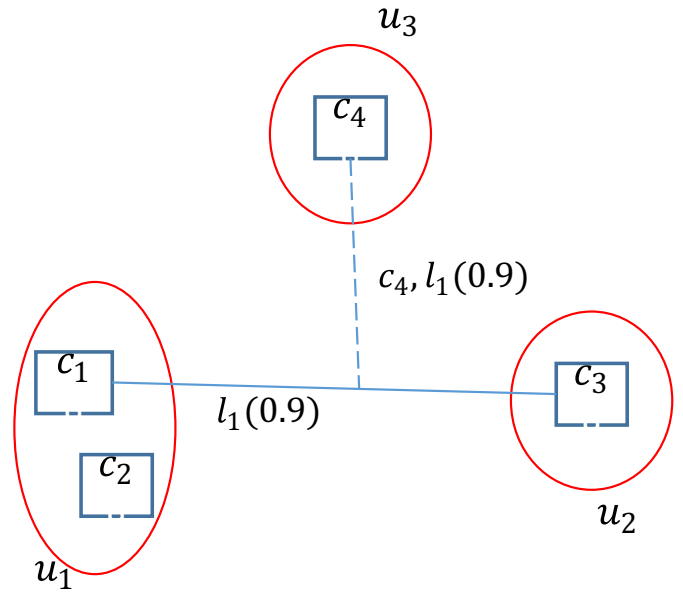
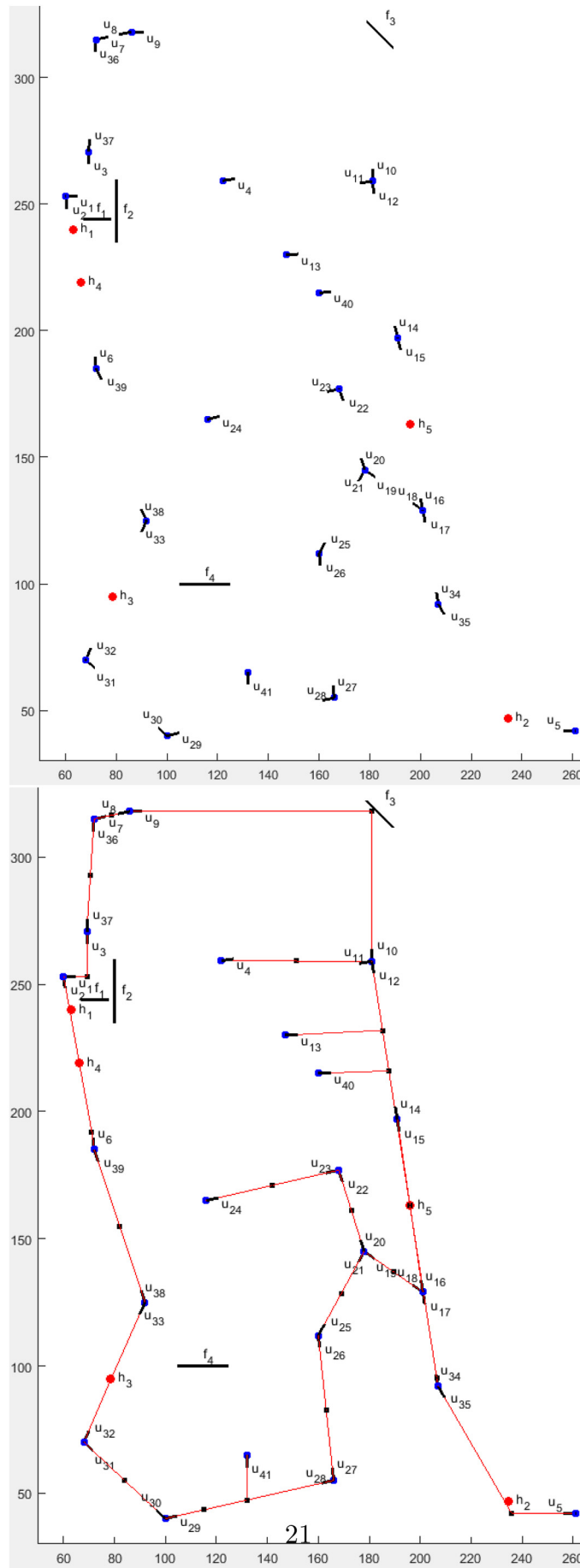


Fig. 10. Optimal solution to the model of the toy example.



**Fig. 11.** The configuration of the synthetic data set (top) including manhole observations  $u_i$ , hypothesized detections  $h_m$ , hypothesized-detection-free measurements  $f_k$ , and at the bottom, the ground truth. The spatial unit is in metres.

$$\begin{aligned}
& \text{maximize } 0.9x_1 + 0.75x_2 + 0.1x_3 + 0.15x_4 + 0.12x_5 \\
& + W(0.01y_{c_1}^{l_5} + 0.02y_{c_2}^{l_5} + 0.1y_{c_3}^{l_3} + 0.2y_{c_3}^{l_4} + 0.9y_{c_4}^{l_1} + 0.5y_{c_4}^{l_2}) \\
& \text{subject to} \quad y_{c_1}^{l_5} \leq x_5, \\
& \quad y_{c_2}^{l_5} \leq x_5, \\
& \quad y_{c_3}^{l_3} \leq x_3, \\
& \quad y_{c_3}^{l_4} \leq x_4, \\
& \quad y_{c_4}^{l_1} \leq x_1, \\
& \quad y_{c_4}^{l_2} \leq x_2, \\
& \quad x_1 + x_2 + x_3 + x_4 + y_{c_1}^{l_5} + y_{c_2}^{l_5} \leq 1, \quad \text{for } u_1 \\
& \quad x_1 + x_2 + x_5 + y_{c_3}^{l_3} + y_{c_3}^{l_4} \leq 1, \quad \text{for } u_2 \\
& \quad x_3 + x_4 + x_5 + y_{c_4}^{l_1} + y_{c_4}^{l_2} \leq 1, \quad \text{for } u_3 \\
& \quad x, y \in \{0, 1\}
\end{aligned}$$

Taking  $W = 0.5$ , the optimal solution of the above model is  $x_1 = 1$ ,  $y_{c_4}^{l_1} = 1$  and all other decision variables are zeros, with the objective value  $0.9 + 0.45 = 1.35$ . Obviously the two most probable connections (both with a probability of 0.9) are chosen. The optimal solution is also depicted in Fig. 10.

## 5. Experimental results

In this section, we present the experimental results with the proposed system on synthetic and real data. We use the  $F_1$  score (aka  $F$ -measure) to assess the quality of our experiment results. We regard each connection pair from the ground truth as a positive result, and each pair that is not linked from the ground truth as a negative result. Let  $TP$ ,  $FP$ ,  $TN$ ,  $FN$  be the numbers of true positive, false positive, true negative and false negative connections respectively. The precision  $Pre$  and recall  $Rec$  are defined as,

$$Pre = \frac{TP}{TP + FP}, \quad Rec = \frac{TP}{TP + FN}, \quad (16)$$

and the  $F_1$  score is calculated as the harmonic mean of both by,

$$F_1 = 2 \times \frac{Pre \times Rec}{Pre + Rec}. \quad (17)$$

The reason for choosing the  $F_1$  score is due to the fact that the distribution of actually connected ( $TP + FN$ ) and unconnected pairs ( $TN + FP$ ) is heavily unbalanced, where we have  $TP + FN \ll TN + FP$ .

In our experiments, the covariance matrix  $\Sigma_n$  in (1) is set as  $\Sigma_n = \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}$ , where the pan and tilt are measured in degrees. The parameter  $b$  used in the Laplace distribution (4) is set as  $b = 0.5$ , which ensures that when  $r_i = r_j$ , the probability density is equal to 1.

### 5.1. Synthetic data

We first apply the proposed system on a synthetic data set. The manhole configuration of this synthetic data set is extracted from the utility records of a real scene. Since there is no depth information in the utility records, the depths of all pipe ends are set to 1. The directions of pipe ends are computed based on the known start points of them and the given intersection points between pipes. Random Gaussian noise of up to 25 degrees is added to the pan and tilt angles of each of them. The locations of hypothesized detections (HDs) are computed by finding the intersection points of CSs with the related pipes and then adding random noise of up to 0.2 m. Several hypothesized-detection-free measurements (HDFs) were also added manually. The observations with respect to this data set are given in Fig. 11 including 22 manholes,

41 utility ends, 5 HDs and 4 HDFs.

Fig. 12 shows two experimental results of using different numbers of HDs and HDFs. In the experiment with respect to the top image in Fig. 12 only 4 HDs and 3 HDFs were used and in the experiment with respect to the bottom image in Fig. 12 all the 5 HDs and 4 HDFs were used. It can be seen that there are more incorrect connections (cyan lines) in the top image since fewer HDs and HDFs were used in the experiment. Compared with the ground truth, the incorrect connections happen at  $u_4$  and  $u_{11}$ . In the ground truth, they are connected by a direct-connection. Because too much noise was added to the direction of  $u_4$  and  $u_{11}$ , in the experiment results, they are wrongly connected to other pipe segments by side-connection. If more HDs are added between  $u_4$  and  $u_{11}$ , this wrong connection can be fixed. In both experiments, the weight parameter  $W$  in Objective (11) is set as 0.41. In different applications, this parameter may need to have other values. As we have mentioned above, to apply the proposed framework on a large area, this parameter should be learned in a smaller area with ground truth.

Table 1 gives a summary of all the results from the three datasets in our experiments in terms of precision, recall and  $F_1$  score, as computed by (16) and (17). It can be observed from the table that precision, recall and  $F_1$  score have all significantly improved after including more HDs and HDFs.

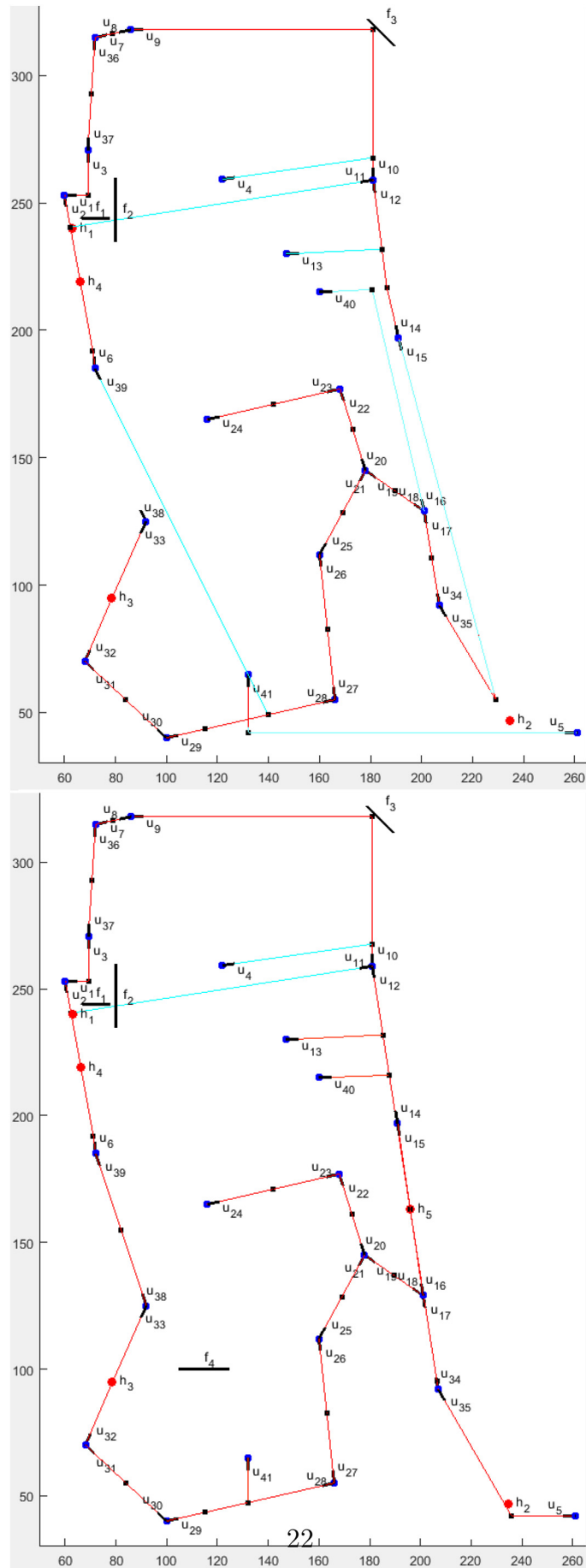
### 5.2. Real data

We also applied the proposed system to a real data set captured on a test scene shown in the left image in Fig. 13. The locations of the manholes were measured by a total station theodolite. The cover of each manhole was lifted and the depth and direction of each utility end observed in the related manhole were measured and estimated. An example manhole picture after the cover was lifted is given by the top right image in Fig. 13 and the related estimated directions of the observed utilities are shown by the bottom right image in Fig. 13. It can be seen that 13 manholes shown by black rectangle windows are observed. 51 utility ends are observed in total from this scene. The utility type and size are also recorded in the observations. There are three different kinds of utilities: foul water, surface water and cable. Since the type and size of the observed utility ends are employed in generating the utility map of the test scene and at the same time different depth values of different utility ends provide more discriminations between different utility segments, the utility map of this scene is generated as expected without using any sensor measurements. In this experiment, the weight parameter  $W$  in Objective (11) is set as 1. It means we do not give any bias to direct-connection. Although we do not have the complete ground truth of this scene a number of pits were dug at carefully chosen locations to verify the correctness of the map; no inaccuracies were discovered. The generated utility map is shown in Fig. 14, and its corresponding precision, recall and  $F_1$  score values are shown in Table 1. In this figure, a top view of this map is shown in the top image and a side view is shown in the bottom image.

### 5.3. ILP solver

After the probability values  $p_i$  and  $p_c^l$  in Eq. (11) were found, the Integer Linear Programming problems were solved on a 64 bit Xpress-MP suite (version 8.0), a commercial optimization package produced by FICO [34], on a Dell workstation with 32G RAM and an Intel Core i7-4790 CPU in Windows 7 Enterprise. The default Integer Linear Programming solver of the suite was used for solving the ILP instances.

All the problem instances were solved within 1 s by the above package.



(caption on next page)

**Fig. 12.** Two experimental results with different numbers of hypothesized detections and hypothesized-detection-free measurements used (the spatial unit is in metres). The red lines indicate correct connections while the cyan lines indicate incorrect connections compared with the ground truth. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**

Accuracy results of all experiments.

Dataset	Precision ( <i>Pre</i> )	recall ( <i>Rec</i> )	$F_1$
Synthetic 1	71.43%	68.18%	69.77%
Synthetic 2	91.30%	95.45%	93.33%
Real-world	100%	100%	100%

## 6. Conclusions

In this paper we introduce a novel method for accurately mapping buried utilities based on the manhole observations from street furniture surveys and sensor measurements. One particular theoretical contribution is the algorithm proposed in Section 3 for calculating the probabilities of the connections under various scenarios. The mapping problem is formulated as an Integer Linear Programming model, which incorporates locations and directions of utility ends, hypothesized sensor detections, hypothesized detection free sensor measurements and the properties of utility ends other than location and direction in the same framework. The experimental results demonstrate the effectiveness and efficiency of the proposed method. In addition, the experimental results on the synthetic data set indicate that the more sensor data is used the more accurate a utility map can be achieved while the experimental results on the real data set demonstrate that using utility properties other than location and direction makes the proposed method more robust and accurate.

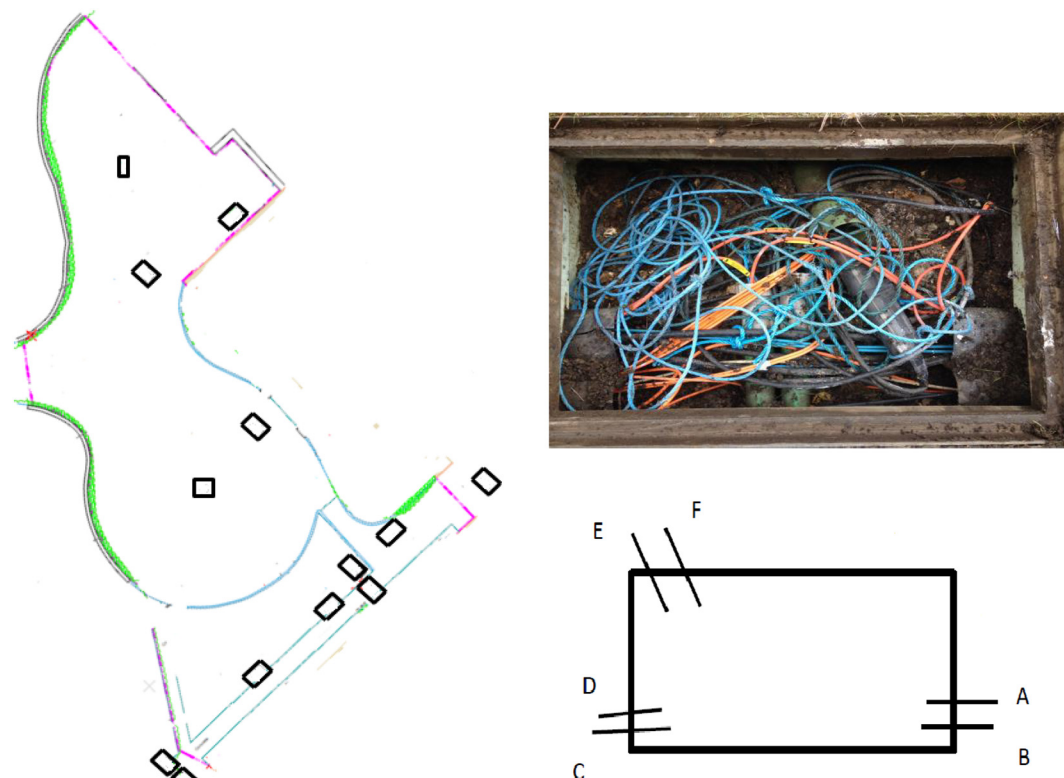
A disadvantage of the proposed framework is that every generated segment has to have at least one utility end observed from a manhole, which may not be always satisfied in practice. We leave for future work

the incorporation of utility segments formed only from sensor measurements into this framework [12]. This should not be difficult for GPR sensor detections since the estimated direction of a utility can be detected from the sensor reading — thus each GPR detection could be regarded as a utility end, rather than an HD. However, an uncertainty would have to be included to represent the possibility that the HD was not in fact of a utility at all (whereas a manhole observation always represents one or more actual utility lines).

Moreover, unfortunately, obtaining real world survey data (including manhole inspections, sensor readings, and verification pits for ground truth) requires a large amount of money which thus limited the size of the real-world instance we could use. We designed the larger sized synthetic instance to better verify our method for larger sized cases. We will leave further verification by large scaled real-world cases to possible opportunities or projects in the future. A third limitation is the need to convert the information provided by the data source into appropriate distributions, such that they can be combined into a compound one.

A future research direction could be how the proposed probabilistic method presented here could be extended to take into account of existing information (e.g. utility records), such as in the work in [1]. This can be realized by converting the existing records into corresponding probability distributions, which is viable and the future work would be about the details of how this conversion could be achieved. Another future direction is to update and improve the quality of existing records. For instance, it might be possible to create a new integer linear programming model to minimize the discrepancies between the manhole/sensor data and the existing records.

Finally, it is worth commenting on the relationship between this



**Fig. 13.** The deployment of manholes (represented by rectangles) in a test scene where the real data has been captured (left image) and an example of the observation in a manhole and the related sketch plan of this manhole.



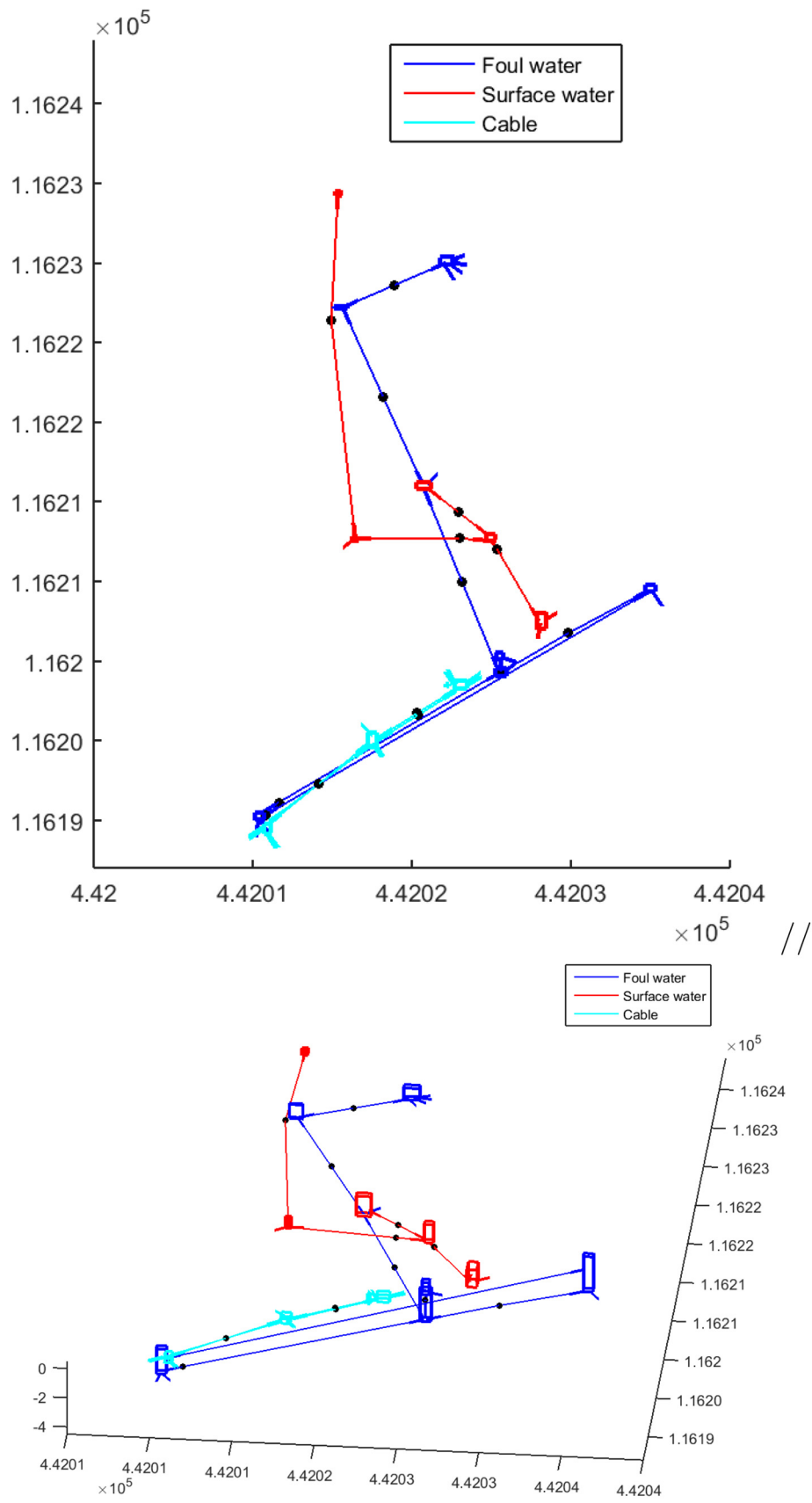


Fig. 14. The generated utility map of the real test scene. The image at the top shows a top view of this map and the bottom image shows a side view (the spatial unit is in metres).

research and the UK PAS 128 standard which is a methodology for delivering robust utility surveys (<https://www.pas128.co.uk>). There are several categories, from D (desktop survey which just uses statutory records, type C (which uses site reconnaissance to match a type D survey with street furniture observations), type B (using multiple geophysical sensors; each 5 m segments should have a quality level for vertical and horizontal detection accuracy) and type A (where position has been verified by trial pits, vacuum extraction etc.). The work presented in this paper would help construct PAS 128 type C surveys in particular, although determining the quality level (QL1–QL4) is not presently supported). The major benefit would be as an automated method for constructing maps from the manhole surveys and sensor measurements. Subsequent manual interpretation would be required to assign quality levels.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

The financial assistance of the EPSRC under grants EP/F06585X and EP/K021699/1 is gratefully acknowledged. The last author is partially supported by a Fellowship from the Alan Turing Institute.

### References

- [1] S. Li, H. Cai, V.R. Kamat, Integrating natural language processing and spatial reasoning for utility compliance checking, *J. Constr. Eng. Manag.* 142 (12) (2016) 04016074, [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001199](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001199).
- [2] J. Cai, S. Li, H. Cai, Accurate mapping of underground utilities: An information fusion approach based on dempster-shafer theory, *Construction Research Congress*, 2018, pp. 712–721, <https://doi.org/10.1061/9780784481295.071>.
- [3] N. Metje, B. Ahmad, S.M. Crossland, Causes, impacts and costs of strikes on buried utility assets, *Proceedings of Institute of Civil Engineers - Municipal Engineer* 168 (3) (2015) 165–174, <https://doi.org/10.1680/muen.14.00035>.
- [4] Pipeline and H. M. S. A. (PHMSA), Significant Pipeline Incidents through 2016 by Cause, url <https://hip.phmsa.dot.gov/analyticssoap/saw.dll?portalpages>, (2017), Accessed date: 11 August 2019.
- [5] T. Bailey, H. Durrant-Whyte, Simultaneous localization and mapping (SLAM): part I, *IEEE Robotics & Automation Magazine* 13.2 (2006) 99–110, <https://doi.org/10.1109/MRA.2006.1638022>.
- [6] T. Bailey, H. Durrant-Whyte, Simultaneous localization and mapping (SLAM): part II, *IEEE Robotics & Automation Magazine* 13 (3) (2006) 108–117, <https://doi.org/10.1109/MRA.2006.1678144>.
- [7] D. Daniels, *Ground Penetrating Radar*, *The Inst. Elect. Eng. (IEE)*, London, U. K., 2004 (ISBN: 0863413609).
- [8] S. Thrun, D. Koller, Z. Ghahramani, H. Durrant-Whyte, A.Y. Ng, *Simultaneous Mapping and Localization with Sparse Extended Information Filters: Theory and Initial Results*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 363–380.
- [9] M. Montemerlo, T. S., Simultaneous localization and mapping with unknown data association using FastSLAM, *Proceedings of 2003 IEEE International Conference on Robotics and Automation*, 2003, pp. 1985–1991, <https://doi.org/10.1109/ROBOT.2003.1241885>.
- [10] M. Bosse, P. Newman, J. Leonard, S. Teller, Simultaneous localization and mapping building in large-scale cyclic environments using the atlas framework, *The International Journal of Robotics Research* 23 (12) (2004) 1113–1140, <https://doi.org/10.1177/0278364904049393>.
- [11] L. Xu, C. Feng, V.R. Kamat, C.C. Menassa, An occupancy grid mapping enhanced visual SLAM for real-time locating applications in indoor GPS-denied environments, *Autom. Constr.* 104 (2019) 230–245, <https://doi.org/10.1016/j.autcon.2019.04.011>.
- [12] Q. Dou, L.J. Wei, D.R. Magee, A.G. Cohn, 3D buried utility location using a marching-cross-section algorithm for multi-sensor data fusion, *Sensors* 16 (11) (2016) 1–24, <https://doi.org/10.3390/s16111827>.
- [13] R. Dutta, A.G. Cohn, J.M. Muggleton, 3D mapping of buried underworld infrastructure using dynamic bayesian network based multi-sensory image data fusion, *J. Appl. Geophys.* 92 (2013) 8–19, <https://doi.org/10.1016/j.jappgeo.2013.02.005>.
- [14] H. Chen, A.G. Cohn, Buried utility pipeline mapping based on multiple spatial data sources: A Bayesian data fusion approach, *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three*, 2011, pp. 2411–2417, <https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-402>.
- [15] R. van Son, S.W. Jaw, J. Yan, V. Khoo, R. Loo, S. Teo, G. Schrotter, A framework for reliable three-dimensional underground utility mapping for urban planning, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4/W10* (2018) 209–214, <https://doi.org/10.5194/isprs-archives-XLII-4-W10-209-2018>.
- [16] J. Yan, S.W. Jaw, W.A. Soon, Kean Huat, G. Schrotter, Towards a 3D Data Model of Underground Utilities for Land Administration, *Preprints*, 2019040223 (2019), <https://doi.org/10.20944/preprints201904.0223.v1>.
- [17] L.L. olde Scholtenhuis, X. den Duijn, S. Zlatanova, Representing geographical uncertainties of utility location data in 3D, *Autom. Constr.* 96 (2018) 483–493, <https://doi.org/10.1016/j.autcon.2018.09.012>.
- [18] F. Döner, R. Thompson, J. Stoter, C. Lemmen, H. Ploeger, P. van Oosterom, S. Zlatanova, 4D cadastres: first analysis of legal, organizational, and technical impact—with a case study on utility networks, *Land Use Policy* 27 (4) (2010) 1068–1081, <https://doi.org/10.1016/j.landusepol.2010.02.003>.
- [19] S. Li, H. Cai, V.R. Kamat, Uncertainty-aware geospatial system for mapping and visualizing underground utilities, *Autom. Constr.* 53 (2015) 105–119, <https://doi.org/10.1016/j.autcon.2015.03.011>.
- [20] S. Li, H. Cai, D.M. Abraham, P. Mao, Estimating features of underground utilities: hybrid GPR/GPS approach, *J. Comput. Civ. Eng.* 30 (1) (2016) 4014108, [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000443](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000443).
- [21] C. Yuan, S. Li, H. Cai, V.R. Kamat, GPR signature detection and decomposition for mapping buried utilities with complex spatial configuration, *J. Comput. Civ. Eng.* 32 (4) (2018) 04018026, [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000764](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000764).
- [22] A.C.D. Royal, P.R. Atkins, M.J. Brennan, Site assessment of multiple-sensor approaches for buried utility detection, *International Journal of Geophysics* 2011 (19) (2011), <https://doi.org/10.1155/2011/496123>.
- [23] J.M. Muggleton, E. Rustighi, Mapping the underworld: recent develops in vibro-acoustic techniques to locate buried infrastructure, *Geotechnique Letters* 3 (3) (2013) 137–141, <https://doi.org/10.1680/geolett.13.00032>.
- [24] Q. Dou, L.J. Wei, D.R. Magee, A.G. Cohn, Real-time hyperbola recognition and fitting in GPR data, *IEEE Trans. Geosci. Remote Sens.* 55 (1) (2017) 51–62, <https://doi.org/10.1109/TGRS.2016.2592679>.
- [25] H. Chen, A.G. Cohn, Probabilistic robust hyperbola mixture model for interpreting ground penetrating radar data, *IEEE World Congress on Computational Intelligence*, 2010, pp. 3367–3374, <https://doi.org/10.1109/IJCNN.2010.5596298>.
- [26] W. Alnuaimy, Y. Huang, M. Nakhkash, M. Fang, V. Nguyen, A. Erisken, Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition, *Journal of Applied Geographics* 43 (2000) 157–165, [https://doi.org/10.1016/S0926-9851\(99\)00055-5](https://doi.org/10.1016/S0926-9851(99)00055-5).
- [27] K. Foo, P. Atkins, A. Thomas, C. Rogers, Capacitive-coupled electric-field sensing for urban sub-surface mapping: Motivations and practical challenges, *International Conference on Frontiers in Shallow Subsurface Technology*, July 2010 <http://www.earthdoc.org/publication/publicationdetails/?publication=37447>.
- [28] P. Wang, K. Goddard, P. Lewin, S. Swingler, Electromagnetic field application to underground power cable detection, *Proceedings of International Symposium on High Voltage Engineering*, (Hannover, Germany), 2011 URI: <https://eprints.soton.ac.uk/id/eprint/272706>.
- [29] J. Muggleton, B.D. Papandreou, M. Brennan, Detection of buried pipes using a shear wave ground surface vibration technique, *Proceedings of 19th International Congress on Sound and Vibration*, July 2012, <https://doi.org/10.1016/j.jappgeo.2014.04.021>.
- [30] Leica, *Leica Flexline ts02/06/09*. Leica Geosystems. See Datasheet for Model TS06 (pdf), <https://leica-geosystems.com/en-gb/products/total-stations/manual-total-stations>, (2008), Accessed date: 11 August 2019.
- [31] A.W. Stroupe, M.C. Martin, T. Balch, Merging Gaussian distributions for object localization in multi-robot systems, *Proceeding of 7th International Symposium on Experimental Robotics (ISER '00)*, December 10–13 2000 <https://link.springer.com/chapter/10.1007/3-540-45118-8>.
- [32] L.A. Wolsey, *Integer Programming*, New York, NY, USA, ISBN-10: 0471283665 Wiley-Interscience, 1998.
- [33] M.R. Garey, D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman & Co., New York, NY, USA, 1979.
- [34] FICO, *Xpress-optimization package*, <http://www.fico.com/en/products/fico-xpress-optimization>, (2017), Accessed date: 26 September 2017.