Gaussian Markov Random Fields for Fusion in Information Form

Liye Sun, Teresa Vidal-Calleja and Jaime Valls Miro

Abstract— 2.5D maps are preferable for representing the environment owing to their compactness. When noisy observations from multiple diverse sensors at different resolutions are available, the problem of 2.5D mapping turns to how to compound the information in an effective and efficient manner. This paper proposes a generic probabilistic framework for fusing efficiently multiple sources of sensor data to generate amendable, high-resolution 2.5D maps. The key idea is to exploit the sparse structure of the information matrix. Gaussian Markov Random Fields are employed to learn a prior map, which uses the conditional independence property between spatial location to obtain a representation of the state with a sparse information matrix. This prior map encoded in information form can then be updated with other sources of sensor data in constant time. Later, mean state vector and variances can be also efficiently recovered using sparse matrices techniques. The proposed approach allows accurate estimation of 2.5D maps at arbitrary resolution, while incorporating sensor noise and spatial dependency in a statistically sound way. We apply the proposed framework to pipe wall thickness mapping and fuse data from two diverse sensors that have different resolutions. Experimental results are compared with three other methods, showing that, while greatly reducing computation time, the proposed framework is able to capture in large extend the spatial correlation to generate equivalent results to the computationally expensive optimal fusion method in covariance form with a Gaussian Process prior.

I. INTRODUCTION

Two-and-a-half dimensional (2.5D) maps have been widely used in robotics to represent the environment in a compact manner. A 2.5D grid map is a two-dimensional grid where each cell stores the measurement at the corresponding place of the environment. For instance, elevation maps represent the height of the surface [1]; occupancy maps store in each cell a binary value which specifies whether a location is occupied or not [2].

In order to build a precise 2.5D grid map, measurements from one or multiple sensors need to be combined in a consistent manner. A key challenge is to exploit the spatial correlation when fusing large amount of data in a theoretically sound and computationally efficient way. While fusion is usually performed within each cell, spatial dependencies between nearby sensor measurements are often ignored. Some recent approaches [3] and [4] consider spatial correlation by using Gaussian Processes (GPs) [5] to fuse large amount of sensor data. However, it requires expensive operations to invert a relatively dense covariance matrix.

This paper proposes a probabilistic framework to integrate measurements from multiple (diverse) sensors into a single high-resolution map. Our previous work [6] avoids inverting

All authors are with the Centre for Autonomous Systems at Faculty of Engineering and IT, University of Technology, Sydney NSW 2007, Australia. `Liye.Sun@student.uts.edu.au`, `Teresa.VidalCalleja@uts.edu.au` and `Jaime.VallsMiro@uts.edu.au`.

a large covariance matrix in Bayesian fusion by building conditionally independent submaps. This work follows the same idea of first building a spatially correlated prior map from one sensor dataset and then updating this prior with another source of sensor data. But we avoid hard cut-off of cross-correlations between submaps and perform Bayesian fusion in information form for the whole map. Based on the recent findings in [7], we approximate the continuous Gaussian Random Fields (GRFs[1]), which is modelled by GPs in [6], using Gaussian Markov Random Field (GMRF) [8]. GMRF also captures the spatial correlation of the data and can infer a prior map. This prior, however, is denoted by the information vector and a sparse information matrix. The sparseness of the information matrix is induced by the Markov property. Bayesian fusion is then performed in information form inexpensively. After fusion, mean and variances can be recovered efficiently by applying sparse matrices techniques.

The contribution of this paper is then two-fold: 1) the use of GMRF to model 2.5D grid maps in information form, which allows efficient inference and more importantly, 2) allows linear time update of the prior map by fusing in information form while considering spatial correlation.

The proposed framework is applied to the scenario of high-resolution pipe wall thickness mapping. For evaluation, we fuse information from two kinds of sensors that are commonly used to assess condition in pipelines. The proposed approach is compared with a fully correlated fusion with an underlying GPs model, a partially correlated GPs model (sparse covariance matrix) [5], and the uncorrelated standard fusion approach. Experimental results demonstrate the validity of the proposed approach in building 2.5D maps at high resolution in an efficient manner.

II. RELATED WORK

In the context of 2.5D grid mapping, there has been extensive work in the literature [1], [2], [9], [10]. In [10], for instance, a statistical representation that accounts for sensing uncertainty of each sensor data is built for a terrain mapping application. Our method is similar to [10] since we also use Bayesian fusion. However, instead of single variances, we consider a prior model to capture the spatial correlation.

GPs, a powerful non-parametric Bayesian technique used for regression and classification, are able to capture the spatial correlation (auto and cross covariance) of data. In recent years, GPs have been introduced for 2.5D mapping. It is demonstrated that GPs could be used to achieve rich and reliable probabilistic inference about unobserved regions [11]. Data fusion is formulated as GPs regression and

---

[1]When the input vectors are 2D, GPs are also known as GRFs.

results are improved by including a covariance matrix. Our previous work [4] and [6] use the prior map generated by GPs for Bayesian fusion. Thus a single measurement influences its neighbourhood by considering the spatial correlations.

Large-scale datasets pose a challenge for Bayesian methods as computation complexity is usually high due to factorising the dense covariance matrix. To tackle this issue, [11] uses KD-tree based local nearest neighbour structure to approximate GPs learning and inference, but the number of neighbours is determined empirically. Other solutions may be found in the GPs approximation methods, such as [12] and [13]. Submapping techniques are also used to account for this problem [6]. Some challenges in submapping include how to define the size of submaps (to what extend cross correlations should be kept) and how to join local submaps properly.

Although expensive in covariance form, Bayesian fusion is "cheap" in information form [14]. Extensive examples in multi-sensor data fusion [15] and Simultaneous Mapping and Localisation (SLAM) [16] show that fusion in information form is more efficient than its dual as only requires an addition operation. However, it is usually expensive to recover the state estimate (mean and variance) unless the information matrix is (enforced to be) sparse. In [16], the sparse information matrix is maintained by using an undirected graph with local relative constraints (edges) between pairs of nearby points in the map, which is in fact a GMRF [17]. The motivation of this paper is then to use a sparse information matrix in Bayesian fusion and thus save computation time, while considering spatial correlations.

GMRFs are discrete domain GRFs equipped with the Markov property. GMRFs enjoy great computational advantages by combining the analytical formulations of Gaussian distributions with the conditional independence property [8]. Although GMRF has not raised much attention in multi-sensor data fusion for 2.5D mapping, an example can be found in [18]. It models local correlations by the numerically defined clique potential functions of neighbourhoods, while the information matrix is not used at all.

III. APPROACH OVERVIEW

We consider information-form Bayesian fusion for 2.5D grid mapping. First, for one sensor dataset, a prior map which contains the spatial correlation is generated using a GMRF. Then Bayesian fusion is done in information form, which takes linear computation time.

Given a sensor dataset $\Psi_1 = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n)\}$ with n point-referenced measurements, where $\mathbf{x}_i \in X$ is the spatial location at which the sensor data $y_i \in Y$ is measured. Ψ_1 is taken from a latent process $y_i = \xi_i + \varepsilon_i$, where $\xi_i = \xi(\mathbf{x}_i) \in \boldsymbol{\xi}$, and $\boldsymbol{\xi}$ is Gaussian distributed. $\varepsilon = \{\varepsilon_i\}_{i=1}^n$ is the noise term. GMRF is used to learn the pdf of $\boldsymbol{\xi}$ from Ψ_1 and we get $p(\boldsymbol{\xi}|X, \Psi_1)$. Data at any points X^* could be inferred by the GMRF by computing $p(\boldsymbol{\xi}|X^*, \Psi_1)$. It is then straightforward to use such conditional distribution as the prior $p(\boldsymbol{\xi}|X) = \mathcal{N}(\boldsymbol{\eta}, Q^{-1})$ for Bayesian fusion. When the new measurements Ψ_2 come, the prior map can be updated. $\Psi_2 = \{(\mathbf{x}_1, z_1), (\mathbf{x}_2, z_2), ..., (\mathbf{x}_k, z_k)\}$, with k sensor readings $\mathbf{z}$ measured from $\mathbf{x}_i \in X^*$ locations. For Ψ_2, a Gaussian model is used as $p(\mathbf{z}|\boldsymbol{\xi}, X) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}}, Q_z^{-1})$, where Q_z^{-1} is the

uncorrelated information matrix. The maximum *a posteriori* estimator is applied to integrate the prior map with the new measurements, and we get the posterior distribution $p(\boldsymbol{\xi}|\mathbf{z}, X) \propto p(\boldsymbol{\xi}|X) p(\mathbf{z}|\boldsymbol{\xi}, X)$.

Finally, the mean state vector and the variances can be efficiently recovered from the information-form as explained in Section V and Section VII.

IV. PRIOR MAP

A. Background: GMRF and Continuously indexed GMRF

GMRFs are MRFs where all variables $\xi_1, \ldots, \xi_n$ are jointly Gaussians. The Markov property is depicted as: for $i \neq j$, ξ_i and ξ_j are independent conditional on the other variables ξ_{-ij}. A GMRF is defined on a set of discretely indexed sites connected by a graph $\mathcal{G}$, which displays conditional independence of $\boldsymbol{\xi}$. $\boldsymbol{\xi}$ is a GMRF with reference to $\mathcal{G}$ with mean $\boldsymbol{\mu}$ and information matrix (inverse of covariance matrix) Q if its pdf is in the below form:

$$p(\boldsymbol{\xi}) \sim \mathcal{N}(\boldsymbol{\mu}, Q^{-1})$$
$$= (2\pi)^{-n/2}|Q|^{1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\xi}-\boldsymbol{\mu})^\top Q(\boldsymbol{\xi}-\boldsymbol{\mu})\right\} \quad (1)$$

where Q encodes the Markov property: $Q_{ij} = 0$ if and only if ξ_i and ξ_j are conditionally independent. A GMRF is specified by Q, and different Q shows different Markov properties of the random field. To construct Q, the particular neighbourhood structure must be defined, according to the data to be modelled.

Recently, [7] proposed a new kind of continuously indexed GMRF, which explicitly maps the Matérn GRF. Assuming $\xi(\mathbf{x})$ is a realisation of random field ξ situated at $\mathbf{x} \in \mathbb{R}^2$, such GMRF is obtained by solving a stochastic partial differential equation (SPDE). Q is then specified, and has an explicit link with the Matérn covariance matrix (see Appendix).

We apply the continuously indexed GMRF to model the spatially correlated data because of two reasons. First, such GMRF is able to capture the structure of spatial correlations and therefore inference can be made in regard of correlations between locations. Second, it does not lose information due to binning the data locations to a regular grid map before applying the GMRF for the grid; and thus the spatial resolution of the grid does not affect the inference significantly.

B. Mapping

The continuously indexed GMRF is used to build the prior map $p(\boldsymbol{\xi}|X) \sim \mathcal{N}(\boldsymbol{\eta}, Q^{-1})$. Considering n observations y_i at locations $\mathbf{x}_i$, $i = 1, \ldots, n$, we define the model as:

$$y_i = \xi_i + \varepsilon_i = f(\mathbf{x}_i) + \varepsilon_i \quad (2)$$
$$\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{m}(X), Q^{-1}(X, X)), \quad (3)$$

where $\mathbf{m}$ is the mean function, f models the underlying process and ε_i is ith noise term. The noise is considered independent, not necessarily identical, distributed as $\mathcal{N}(0, \sigma_{\varepsilon,i}^2)$. When the noise is assumed to be constant, $\sigma_{\varepsilon,i}^2$ can be learned together with other parameters. The GMRF model is specified by mean function m and parameters in Matérn covariance function (denoted as $\boldsymbol{\theta}$). Without loss of generality zero mean function for $\mathbf{m}$ can be assumed. The

parameters $\boldsymbol{\theta}$ are then learned by solving SPDE using finite element method [7] and used to build the information matrix Q, as is explained in the Appendix.

Predicting the information matrix and the mean vector at the query locations X^* is done by using:

$$Q = Q(X^*, X^*) + Q_\varepsilon \tag{4a}$$

$$\boldsymbol{\mu} = \boldsymbol{m}(X^*) + Q^{-1}Q(X^*, X)(\boldsymbol{y} - \boldsymbol{m}(X)). \tag{4b}$$

Here Q_ε is a diagonal matrix that represents the precision of data itself, i.e., the inverse of variance of data noise. In our particular case, we do not consider the sensor noise to be constant. Note that with the zero-mean function, the information vector $\boldsymbol{\eta}$ corresponding to $\boldsymbol{\mu}$ becomes:

$$\begin{aligned} \boldsymbol{\eta} &= Q\boldsymbol{\mu} \\ &= Q(X^*, X)\boldsymbol{y}. \end{aligned} \tag{5}$$

The comparison between (4b) and (5) shows that maintaining $\boldsymbol{\eta}$ is cheaper than estimating its dual. Therefore we use $\boldsymbol{\eta}$ for fusion. When $\boldsymbol{\mu}$ is needed, the expensive calculation of $Q^{-1}Q(X^*, X)(\boldsymbol{y} - \boldsymbol{m}(X))$ in (4b) can be done by first computing sparse Cholesky factorisation $Q = LL^\top$ and then solving $Qx = Q(X^*, X)(\boldsymbol{y} - \boldsymbol{m}(X))$ using forward and backward substitutions. Reordering methods (nested dissection [19], CAMD [20], etc.) are often required to achieve sparsity. For large-scale data ($n > 10^6$) iterative equation solvers [21] can provide an approximate solution.

V. FUSION IN INFORMATION FORM

Having obtained the spatially correlated prior map $p(\boldsymbol{\xi}|X) \sim \mathcal{N}(\boldsymbol{\eta}, Q^{-1})$ from Section IV-B, we model the new independent measurements $\mathbf{z}$ as Gaussian distributed $p(\mathbf{z}|\boldsymbol{\xi}, \mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}}, Q_z^{-1})$ and use it to update the prior map by the maximum *a posteriori* estimator. We obtain the posterior density $p(\boldsymbol{\xi}|\mathbf{z}, X)$ in information form as,

$$\boldsymbol{\eta}^+ = \boldsymbol{\eta} + H^\top Q_z \mathbf{z} \tag{6a}$$

$$Q^+ = Q + H^\top Q_z H, \tag{6b}$$

where H is the observation matrix, which selects part of state $\xi_1, \ldots, \xi_n$ that is observed by $\mathbf{z}$. H equals to $I_{n\times n}$ when the full state $\boldsymbol{\xi}$ is observed through $\mathbf{z}$. Q_z is a diagonal matrix with the entries being the inverse of constant or non-constant noise variances $\sigma_{z,i}^2$. By regarding the fusion results as new prior, any new data sources can update the map.

Note that even the most expensive computation $H^\top Q_z H$ in (6) is done in constant time. The update acts to strengthen existing constraints between measurements. In (6b), the addition of $H^\top Q_z H$ does not make denser the prior information matrix Q. The posterior mean $\boldsymbol{\mu}^+$ can be recovered by solving the sparse linear system $Q^+x = \boldsymbol{\eta}^+$ in a similar way as in Section IV-B. The posterior variances could also be recovered efficiently using the approach proposed by [22].

VI. APPLICATION TO THICKNESS MAPPING

A. Sensor information

The proposed approach is applied to 2.5D thickness mapping for water pipes. We fuse two data sources that contain pipe's remaining wall thickness measurements: 1)

a pulsed-Eddy current sensor [23] (*abbr.* LR sensor) and 2) a magnetic flux leakage sensor [24] (*abbr.* HR sensor) both modelled with non-constant noise. Information of these two heterogeneous sensors correspond to data-sets Ψ_1 and Ψ_2, respectively. Another data source generated using a 3D laser scanner [25] is used as ground-truth.

All the sensor measurements were taken from a real cast-iron pipe section of 1000 mm length and 2073.5 mm circumference. HR sensor and LR sensor cover the full area with a 2D grid of 192×100 and 42×20 respectively. The ground-truth covers the full circumference and the 0 to 891.8 mm length, and each cell's size is 1.2mm × 1.2mm. Therefore quantitative comparisons with ground-truth are done from 0 to 891.8 mm length and the whole circumference.

LR sensor data (mean thickness map and uncertainty map) are shown in Fig. 1(a) and Fig. 2(a). HR sensor data are shown in Fig. 1(b) and Fig. 2(b). HR sensor data are very sparse and contain only 2043 readings. Neither LR nor HR sensor has constant noise, see Fig. 2(a) and Fig. 2(b). The ground-truth is shown in Fig. 1(c). Note that in these 2.5D plots, thickness and uncertainty are shown in colour. Also, the Cylindrical coordinates of the real pipe have been converted into Cartesian coordinates, *i.e.*, vertical axis corresponds to the circumferential axis of the pipe and the horizontal axis is the longitudinal axis.

B. Prior thickness map

Either dataset could be used to train the GMRF model. GMRF will assign the gaps in the sensor data the probabilistic values that are correlated with neighbouring areas covered by the sensors. The continuously indexed GMRF is able to increase/decrease the resolution of the grid map by inference. Even with data observed at irregular grids, the piecewise linear approximation method [7] could model the spatial correlation properly, whereas the traditional GMRF model can only handle X^* with the same resolution as X appropriately. Aiming at predicting the high-resolution map, we use LR sensor dataset to learn the continuously indexed GMRF and conduct fusion at the same resolution as the HR sensor dataset (vice-versa is possible).

We use a zero mean function and Matérn covariance function in GMRF model. The mean and covariance function could be set to be others with some modification to the SPDE model [7]. The order of Matérn covariance function is set to be $\nu = 2$. Other parameters are learned via solving the SPDE. Our implementation uses the R-INLA package [26]. R-INLA first approximates $\boldsymbol{\xi}$ with piecewise linear basis functions, which are defined on a triangular domain, and then gets the GMRF by solving (12) using finite element methods (FEM). This approach is proved to be the best linear approximation to the continuous solution to SPDE [7]. Note that the parameters of SPDE have a direct link with that of Matérn covariance function (see Appendix).

Then probabilistic thickness map of LR sensor at 192×100 resolution (see Fig. 3(a) and Fig. 3(d)) is obtained by the sum of linear predictor components. However, this approach cannot produce the conditional information matrix for the queried locations, but only produces the information matrix of triangular basis functions. To obtain the desired
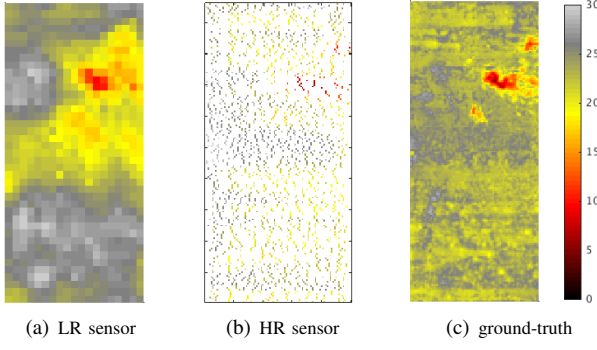
(a) LR sensor   (b) HR sensor   (c) ground-truth

Fig. 1: Thickness maps (mean) from available datasets.

(a) LR sensor   (b) HR sensor

Fig. 2: 2σ uncertainty maps from available datasets.

information matrix, we first calculate the coefficients of information matrix using the *Result 1* of [7], and then optimise the parameters of the information matrix by minimising its difference with the real Matérn covariance matrix.

C. Bayesian fusion in information form

In Bayesian fusion, the prior map will be updated using the HR sensor data. In information form, the update can be done cheaply. After obtaining the information-form posterior map, the mean estimates and the variances can be recovered very efficiently following the methods in Section V. By fusion, the low resolution, but dense data from LR sensor and high resolution, but sparse data from the HR sensor are integrated into a dense and high-resolution map.

VII. EXPERIMENTAL RESULTS

This section compares the fusion results of the proposed method and three other comparison methods.[2] We first briefly describe these three methods as follows.

Method 1 (optimal) uses GPs, with zero mean function and Matérn covariance function, to learn the spatially correlated data from LR sensor, and then infer the high-resolution map $p(\boldsymbol{\xi}|X) \sim \mathcal{N}(\boldsymbol{\mu},\Sigma)$, which is used as the prior map for fusion. The HR sensor readings are modelled as $p(\mathbf{z}|\boldsymbol{\xi},X) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}},R)$, where $R = Q_z^{-1}$ is a diagonal matrix with sensor noise variances entries. The prior map is updated with HR sensor data by Bayesian fusion via maximum *a posteriori*, and the posterior map is:

$$\boldsymbol{\mu}^+ = \boldsymbol{\mu} + \Sigma H^\top (H\Sigma H^\top + R)^{-1}(\mathbf{z} - H\boldsymbol{\mu}) \quad (7)$$

$$\Sigma^+ = \Sigma - \Sigma H^\top (H\Sigma H^\top + R)^{-1}H\Sigma. \quad (8)$$

Method 1 is able to cover the long-length correlation, and is optimal in the sense that the full correlation is modelled. However, there are many fill-ins in the covariance matrix.

Method 2 is almost the same as Method 1, with the only difference that the compactly supported piecewise polynomial covariance function $K_{pp,2}(r)$ [5] is used in GPs. We denote such GPs as *sparse GPs*.

$$K_{pp,2}(r) = \sigma^2 \max(1-r,0)^{j+2} \big((j^2 + 4j + 3)\, r^2 + (3j+6)\, r + 3 \big)/3, \quad (9)$$
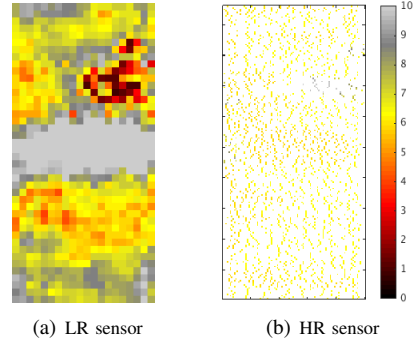
where r is the distance between positions; σ^2 is the variances. $j = \lfloor D/2 \rfloor + 3$ and D is the data dimension, here $D = 2$. Such covariance function generates a sparse covariance matrix by forcing correlations between faraway cells to be zero.

Method 2 is similar to our approach in the sense that only correlations between adjacent points are kept. However, the compactly supported kernel cuts the correlation in a hard way while our approach learns the best approximation. In fairness, we choose parameters in (9) in such a way that the cross-correlation of the sparse covariance matrix stays 0.1, which covers the same correlation distance as ours.

Method 3 is standard Bayesian fusion using the uncorrelated measurements, denoted as *Naïve fusion*. The spatial correlation from the LR sensor data is not learned. The posterior mean and variances are:

$$\mu_i^+ = \big(\sigma_{\varepsilon,i}^{-2} y_i + \sigma_{z,i}^{-2} z_i\big) \big/ \big(\sigma_{\varepsilon,i}^{-i} + \sigma_{z,i}^{-2}\big) \quad (10)$$

$$\sigma_i^{2+} = 1 \big/ \big(\sigma_{\varepsilon,i}^{-i} + \sigma_{z,i}^{-2}\big). \quad (11)$$

The computation complexity of Method 3 is linear. However, point measurement is not propagated to the neighbouring points since there are no correlations between them.

The prior map inferred by the proposed method, Method 1 and 2 are shown in Fig. 3. Note that since Method 3 cannot infer the prior map, we use our prior map in Method 3 for the fusion. The posterior maps generated by the proposed method and Method 1, 2 and 3 are shown in Fig. 4. It shows that the proposed method generates fusion results similar to that of Method 1, which is optimal. Since Method 2 only considers local correlation, one single measurement only updates a small number of neighbours. Method 3 does not include spatial correlation and point-to-point fusion is done. It is obvious that all fusion methods reduce the uncertainty. Since GMRF is conservative, the uncertainty of the proposed approach is larger than that of Method 1. Since correlations between far-away locations are ignored, either in a static way as Method 2 and Method 3, or in a continuous way as the proposed method, the uncertainty map produced by these methods are not as 'smooth' as the fully correlated fusion.

For quantitative evaluation, we compute the root mean squared error (*rmse* in *mm*) for the mean thickness maps using the down-sampled laser scanning data as ground-truth. The *rmse* of LR sensor data and HR sensor data (Fig. 1) are 3.45 and 2.69, respectively. The *rmse* of the inference (Fig.

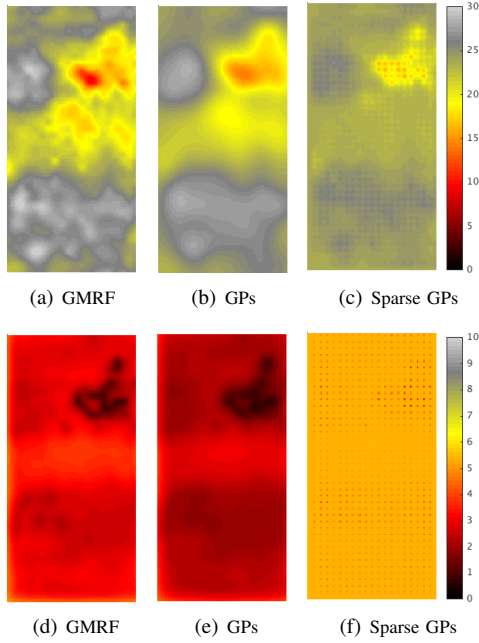[2]For simplicity, we denote these three methods as Method 1, Method 2 and Method 3 in the following paper.

Fig. 3: Inferred thickness maps at high resolution. Mean values (top) and 2σ values (bottom).

TABLE I: General comparison of computation cost.

	Method	CPU Time(s)	Computation Complexity
Our method	GMRF inference	0.15	$O(N^{3/2})$ [a]
	Information-form Bayesian fusion	0.04	$O(N)$
	Recovery of mean and variances	9.47	depends on real problem [b]
Method 1	GPs inference	13.59	$O(N^3)$
	Covariance-form Bayesian fusion	124.30	$O(N^3)$
Method 2	Sparse GPs inference	12.12	$O(MN^2)$ [c]
	Covariance-form Bayesian fusion	111.45	$O(N^3)$
Method 3	No inference	N/A	N/A
	Naïve fusion using our prior map	0.04	$O(N)$

[a] With a 2D GMRF model and computations are done with information matrix. Reordering methods are usually needed.
[b] Depends on the sparsity of information matrices. Here it is $O(N)$.
[c] M is the number of non-zero entries in the covariance matrix, $M < N$.

3) are $(a):3.41$, $(b):3.24$, $(c):3.33$. The *rmse* for each picture in Fig. 4 is $(a):1.98$, $(b):1.91$, $(c):2.79$, $(d):3.21$.

Table I compares the computation time and complexity. All computations are done on a workstation with sixteen 3.10 Ghz Intel Xeon E5-2687W processors. GPs modelling and Bayesian fusion in covariance form are implemented in Matlab. GMRF computations and Bayesian fusion in information form are done in R Studio. Table I shows that both GPs inference and Bayesian fusion suffer the cubic operation while the computations with the sparse information matrix is much cheaper. When the sparsity of the information matrix is changed (or the correlations between points), the complexity shown in Table I remains the same for all rows except for the cost of recovering the mean and variance as it will increase when data become more correlated. For an information matrix with a general sparse structure, reordering will is done first and then Choelsky factorisation costs $O(N^{3/2})$, producing a substantial complexity reduction when N is large. Another observation from Table I is that fusion in Method 2 costs almost the same time as Method 1. This is because the GPs inference destroys the zero pattern of the sparse covariance matrix and the covariance matrix gets populated. We also compare the complexity for fusion when all the methods are formulated in information form. Generally speaking, Methods 1 and 2 require $O(N^2)$ for information-form fusion and $O(N^3)$ to recover the mean and variances. However in the latter, complexity can often be reduced by exploiting information matrix sparsity. In our application, the information matrix recovered from the GPs is less sparse than the one obtained by the GMRF.

VIII. CONCLUSION

This paper proposes a generic framework to fuse multiple sources of (different) sensor data into a probabilistic 2.5D grid map in a computationally efficient way. This framework is capable of handling large-scale data and generating high-resolution maps, while modelling (heterogeneous) sensor noise and spatial correlation in a statistically sound way.

A novel GMRF representation is first built on one source of spatially correlated sensor data. Such GMRF model is an explicit mapping to the latent continuous field and is able to infer the high-resolution map from low-resolution data efficiently. The predicted high resolution map, represented in information form, is then used as the prior for Bayesian fusion and is updated with another source of sensor data. Bayesian fusion with the sparse information matrix allows to update the whole map with the sparse observations within constant time. The mean and uncertainty map are then recovered in constant time by exploiting the sparseness structure.

The proposed work is applied to the 2.5D mapping of pipe's remaining wall thickness. Experimental results, when compared with three standard fusion methods, demonstrate the validity of the proposed work in generating accurate thickness maps in a highly efficient manner.

ACKNOWLEDGMENT

REFERENCES

[1] S. Lacroix, I.-K. Jung, and A. Mallet, "Digital elevation map building with low altitude stereo imagery," *Robot. Auton. Syst.*, vol. 41, no. 2, pp. 119–127, 2002.
[2] H. Moravec and A. E. Elfes, "High resolution maps from wide angle sonar," in *Proceedings of the IEEE International Conference on Robotics and Automation*, March 1985, pp. 116 – 121.
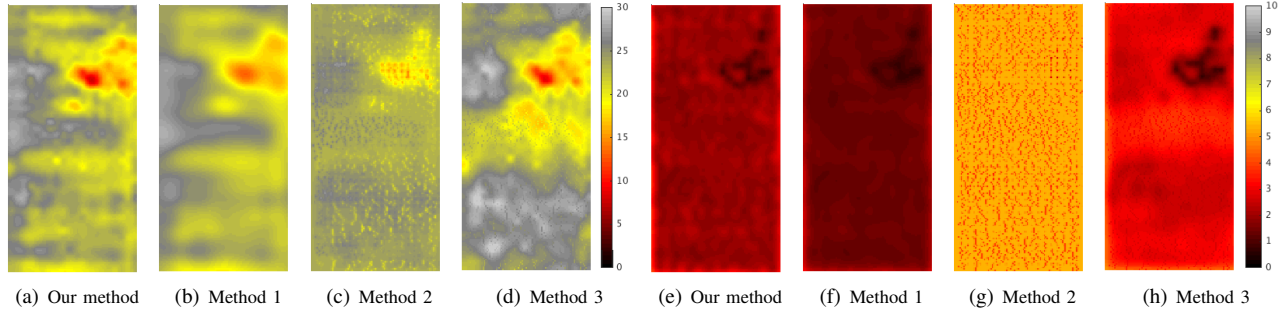
| (a) Our method | (b) Method 1 | (c) Method 2 | (d) Method 3 | (e) Our method | (f) Method 1 | (g) Method 2 | (h) Method 3 |

Fig. 4: Fusion results: thickness maps (mean) at high resolution ((a) to (d)) and 2σ uncertainty maps ((e) to (h)). Note that we use our prior map for fusion in Method 3.

[3] S. Vasudevan, F. Ramos, E. Nettleton, and H. Durrant-Whyte, "Gaussian process modeling of large-scale terrain," *J. Field Robot.*, vol. 26, no. 10, pp. 812–840, 2009.

[4] T. Vidal-Calleja, D. Su, F. De Bruijn, and J. V. Miro, "Learning spatial correlations for Bayesian fusion in pipe thickness mapping," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2014, pp. 683–690.

[5] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for machine learning*. Cambridge, Mass.: MIT Press, 2006.

[6] L. Sun, T. Vidal-Calleja, and J. Valls Miro, "Bayesian fusion using conditionally independent submaps for high reolution 2.5d mapping," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2015, pp. 3394–3400.

[7] F. Lindgren, H. Rue, and J. Lindstrm, "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 73, no. 4, pp. 423–498, 2011.

[8] H. Rue and L. Held, *Gaussian Markov random fields: theory and applications*. CRC Press, 2005.

[9] M. Montemerlo and S. Thrun, "A multi-resolution pyramid for outdoor robot terrain perception," in *AAAI*, vol. 4, 2004, pp. 464–469.

[10] I. Miller and M. Campbell, "A mixture-model based algorithm for real-time terrain estimation," *J. Field Robot.*, vol. 23, no. 9, pp. 755–775, 2006.

[11] S. Vasudevan, F. Ramos, E. Nettleton, and H. Durrant-Whyte, "Heteroscedastic Gaussian Processes for data fusion in large scale terrain modeling," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2010, pp. 3452–3459.

[12] J. Quiñonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian Process regression," *J. Mach. Learn. Res.*, vol. 6, pp. 1939–1959, 2005.

[13] N. Cressie and G. Johannesson, "Fixed rank kriging for very large spatial data sets," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 70, no. 1, pp. 209–226, 2008.

[14] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[15] H. Durrant-Whyte and T. Henderson, *Multisensor Data Fusion*. Springer Berlin Heidelberg, 2008, book section 26, pp. 585–610.

[16] M. R. Walter, R. M. Eustice, and J. J. Leonard, "Exactly sparse extended information filters for feature-based slam," *Int. J. Robot. Res.*, vol. 26, no. 4, pp. 335–359, 2007.

[17] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Comput.*, vol. 13, no. 10, pp. 2173–2200, 2001.

[18] R. Tse, N. R. Ahmed, and M. Campbell, "Unified terrain mapping model with markov random fields," *IEEE Trans. Robot.*, vol. 31, no. 2, pp. 290–306, 2015.

[19] A. George, "Nested dissection of a regular finite element mesh," *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 345–363, 1973.

[20] Y. Chen, T. A. Davis, W. W. Hager, and S. Rajamanickam, "Algorithm 887: Cholmod, supernodal sparse cholesky factorization and update/downdate," *ACM Trans. Math. Softw.*, vol. 35, no. 3, p. 22, 2008.

[21] Y. Saad, *Iterative methods for sparse linear systems*. Siam, 2003.

[22] G. H. Golub and R. J. Plemmons, "Large-scale geodetic least-squares adjustment by dissection and orthogonal decomposition," *Linear Algebra Appl.*, vol. 34, pp. 3–28, 1980.

[23] N. Ulapane, A. Alempijevic, T. Vidal-Calleja, J. V. Miro, J. Rudd, and M. Roubal, "Gaussian Process for interpreting pulsed eddy current signals for ferromagnetic pipe profiling," in *IEEE Int. Conf. Ind. Elect. Appl. (ICIEA)*, 2014, pp. 1762–1767.

[24] B. Wijerathna, T. Vidal-Calleja, S. Kodagoda, Q. Zhang, and J. V. Miro, "Multiple defect interpretation based on Gaussian Processes for MFL technology," *SPIE Smart Struct. Mater. Nondestruct. Charact. Compos. Mater., Aerosp. Eng., Civ. Infrastruct., Home. Secur.*, 2013.

[25] B. Skinner, T. Vidal-Calleja, J. V. Miro, F. De Bruijn, and R. Falque, "3D point cloud upsampling for accurate reconstruction of dense 2.5 D thickness maps," in *Australas. Conf. Robot. Autom. (ACRA)*, 2014.

[26] F. Lindgren and H. Rue, "Bayesian spatial modelling with R-INLA," *J. Stat. Softw.*, vol. 63, no. 19, pp. 1–25, 2015.

APPENDIX
GMRF APPROXIMATION TO MATÉRN GRF

This appendix explains some background knowledge about the continuously indexed GMRF [7] at two dimension.

Assuming $\xi(\boldsymbol{x})$ is a realisation of random field ξ situated at $\boldsymbol{x} \in \mathbb{R}^2$, the continuously indexed GMRF is obtained by solving a stochastic partial differential equation (SPDE):

$$(\kappa^2 - \Delta)^{\alpha/2}[\tau\xi(\boldsymbol{x})] = W(\boldsymbol{x}), \quad \kappa > 0, \quad \tau > 0. \tag{12}$$

Here κ is the spatial scale. Δ is the 2D Laplacian operator. α controls the smoothness of the realisation. τ controls the variances. W is the Gaussian white noise. These parameters are explicitly linked to the parameters of Matérn covariance function, which is:

$$K_{Matern}(X, X^*) = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa d)^\nu K_\nu(\kappa d) \tag{13}$$

where σ_f represents the variance. Γ is the Gamma function. d denotes the distance between input locations. In our pipe thickness mapping task, d is the shortest distance between two points on the cylinder surface. $K_\nu(.)$ is the modified Bessel function of the second kind of order ν. The number of neighbouring cells that are conditioned on depends on ν. The range parameter l controls the distance at which the correlation decreases almost to null. The link between the SPDE solution and Matérn covariances is:

$$\sigma_f^2 = \Gamma(\nu)(\Gamma(\alpha)4\pi\kappa^{2\nu}\tau^2)^{-1} \tag{14}$$

$$\nu = \alpha - 1 \tag{15}$$

$$l = \sqrt{8\nu}/\kappa \tag{16}$$

Typically, (16) is the empirical definition of l, which makes the Matérn correlation stay fairly around 0.1 at a distance l.

When $\nu = 0$, in the continuously indexed GMRF, the information matrix for one single site is

$$\begin{array}{ccc} & -1 & \\ -1 & a & -1 \\ & -1 & \end{array} \tag{17}$$

where $a = \kappa^2 + 4$. [7] derived that the coefficients of information are computed by convolving (17) ν times.