2009 IEEE International Conference on Robotics and Automation
Kobe International Conference Center
Kobe, Japan, May 12-17, 2009

Gaussian Process Modeling of Large Scale Terrain

Shrihari Vasudevan, Fabio Ramos, Eric Nettleton and Hugh Durrant-Whyte
Australian Centre for Field Robotics (ACFR)
The University of Sydney, NSW 2006, Australia
{s.vasudevan,f.ramos,e.nettleton,hugh}@acfr.usyd.edu.au

Allan Blair
Technology and Innovation
Rio Tinto
allan.blair@riotinto.com

Abstract— **This paper addresses the problem of large scale terrain modeling for a mobile robot. Building a model of large scale terrain data that can adequately handle uncertainty and incompleteness in a statistically sound way is a very challenging problem. This work proposes the use of Gaussian Processes as models of large scale terrain. The proposed model naturally provides a multi-resolution representation of space, incorporates and handles uncertainties aptly and copes with incompleteness of sensory information. Gaussian Process Regression techniques are applied to estimate and interpolate (to fill gaps in unknown areas) elevation information across the field. The estimates obtained are the best linear unbiased estimates for the data under consideration. A single Non-Stationary (Neural Network) Gaussian Process is shown to be powerful enough to model large and complex terrain, handling issues relating to discontinuous data effectively. A local approximation methodology based on KD-Trees is also proposed in order to ensure local smoothness and yet preserve the characteristic features of rich and complex terrain data. The use of the local approximation technique based on KD-Trees further addresses concerns relating to the scalability of the proposed approach for large data sets. Experiments performed on sparse GPS based survey data as well as dense laser scanner data taken at different mine-sites are reported in support of these claims.**

I. Introduction

Large scale terrain mapping is a difficult problem with wide-ranging applications, from space exploration to mining and more. For autonomous robots to function in such high-value applications, an efficient, flexible and high-fidelity representation of space is critical. The key challenges posed by this problem are that of dealing with the problems of uncertainty, incompleteness and handling unstructured (potentially highly) terrain. Uncertainty and incompleteness are virtually ubiquitous in robotics as sensor capabilities are limited. The problem is magnified in a field robotics scenario due to sheer scale of the application (for instance, a mining or space exploration scenario).

State-of-the-art representations generally map surfaces by computing triangulations. This process, however, does not have a statistically sound way of incorporating and managing uncertainty. The assumption of statistically independent data is a further limitation of many works that have used these approaches. While there are several interpolation techniques known, the independence assumption can lead to simplistic (simple averaging like) techniques that result in an inaccurate modeling of the terrain. Further, the limited perceptual capabilities of sensors renders most sensory data incomplete. These problems are addressed in this paper by putting forward a Gaussian Process based representation of large scale terrain. The method incorporates sensor uncertainty and uses it for learning the terrain model. Estimation of data at unknown locations is treated as a regression problem in 3D space and takes into account the correlated nature of the spatial data. This technique is also used to overcome sensor limitations and occlusions by filling the gaps in sensor data with the best linear unbiased estimates. The representation is a continuous domain, compact and non-parametric model of the terrain and hence can readily be used to create terrain maps at any required resolution.

The contribution of this paper is a novel approach towards representing large scale terrain using Gaussian Processes. Specifically, this paper shows that a single non-stationary kernel (neural-network) Gaussian Process is successfully able to model large scale terrain data taking into account the local smoothness as well as preserving much of the spatial features in the terrain. A further contribution to this effect is a local approximation methodology based on KD-Trees that incorporates the benefits of both stationary and non-stationary kernels in the elevation estimation process. The use of the KD-Tree based local approximation technique also enables this work to take into account scalability considerations for handling large scale, complex terrain. The paper also compares the performance of stationary and non-stationary kernels. The end-result is a multi-resolution representation of space that incorporates and manages uncertainty in a statistically sound way, handling spatially correlated data in an appropriate manner. Experiments conducted on real sensor data obtained from GPS and Laser scanner based surveys in real application scenarios (mining) clearly suggest the viability of the proposed technique.

II. Related Work

State-of-the-art representations used in applications such as mining, space exploration and other field robotics scenarios as well as in geospatial engineering are typically limited to elevation maps, triangulated irregular networks (TIN's), contour models and their variants or combinations ([1] and [2]). Each of them have their own strengths and preferred application domains. The former two are more popular in robotics. The latter one represents the terrain as a succession of "isolines" of specific elevation (from minimum to maximum). They are particularly suited for modeling hydrological phenomena and otherwise offer no particular computational advantages for the context of this paper.

Grid based methods represent space in terms of elevation data corresponding to each cell of a regularly spaced grid structure. The outcome is a 2.5D representation of space. The main advantage of this representation is simplicity. The main limitations include the inability to handle abrupt changes, the dependence on grid size and the issue of scalability in large environments. In robotics, grid based methods have been exemplified by numerous works such as [3], [4], [5] and more recently [6]. Both [3] and [4] use laser range finders to construct elevation maps. The latter also proposes a "certainty assisted spatial filter" that uses "certainty" information as well as spatial information (elevation map), to filter out erroneous terrain pixels. The authors of [5] developed stereo-vision based elevation maps. They recognize that the main problem is uncertainty incorporation and management. They proposed a heuristic data fusion algorithm that was based on the Dempster Shafer theory. Triebel et al in [6] propose an extension to standard elevation maps in order to handle multiple surfaces and overhanging objects. The main weakness observed in most prior work in grid based representations is the lack of a statistically direct way of incorporating and managing uncertainty.

Triangulated Irregular Networks (TIN) usually sample a set of surface specific points that capture all important aspects of the surface to be modeled - bumps/peaks, troughs, breaks etc. The representation typically takes the form of an irregular network of such points with each point linked to its immediate neighbors. This set of points is represented as a triangulated surface. TIN's are able to more easily capture sudden elevation changes and are also more flexible and efficient than grid maps in relatively flat areas. In robotics, TIN's have been used in works such as [7] and [8]. [7] presents an approach to performing online 3D multi-resolution reconstruction of unknown and unstructured environments to yield a stochastic representation of space. [8] is a recent work in the space exploration domain, that presents a LIDAR based approach to terrain modeling using TIN's. The work further decimates co-planar triangles into a single triangle to provide a more compact representation. TIN's may be efficient from a survey perspective as few points are hand-picked. However [6] points out that for dense sensor data, while they are accurate and can easily be textured, they have a huge memory requirement which grows linearly with the number of scans. The assumption of statistical independence may render the method ineffective at handling incomplete data as the elemental facet of the TIN (planar triangle) may approximate a complicated surface beyond acceptable limits. This however depends on the choice of the sensor and the data density obtained.

There are several different kinds of interpolation strategies for grid data structures. The choice of the interpolation method can have severe consequences on the accuracy of the model obtained. Kidner ([9]) reviewed and compared numerous interpolation methods for grid based methods. He recommended the use of higher order polynomial interpolation (at least Bicubic) methods as a requirement for grid data interpolation. Ye and Borenstein in [4] used median filtering to fill in the missing data in an elevation map. The Locus algorithm [10] was used in [3]. The interpolation method attempted to find an elevation estimate by computing the intersection of the terrain with the vertical line at the point in the grid - this was done in image space rather than Cartesian space.

Gaussian Processes [11] (GP's) are powerful non-parametric learning techniques that can handle all problems mentioned above. They produce a scalable multi-resolution model of the large scale terrain under consideration. This is due to the fact that they yield a continuous domain representation of the terrain data and hence can be sampled to meet any desired resolution easily. They incorporate and handle uncertainty in a statistically sound way and represent spatially correlated data in an appropriate manner. They model and use the spatial correlation of the given data points to estimate the elevation values for other unknown points of interest. In an estimation sense, GP's provide the best linear unbiased estimate [12] based on the underlying stochastic model of the spatial correlation between the data points. They basically perform an interpolation methodology called *Kriging* [13] which is a standard interpolation technique used in the mining industry. Hence, GP's handle both uncertainty and incompleteness effectively.

In the recent past, Gaussian Processes have been applied in the context of terrain modeling - [14], [15] and [16]. All three works are based on using a non-stationary equivalent of a stationary squared exponential covariance function [17] and incorporate kernel adaptation techniques in order to adequately handle both smooth surfaces as well as inherent (and characteristic) surface discontinuities. Whereas [14] initializes the kernel matrices evaluated at each point with parameters learnt for the corresponding stationary kernel and then iteratively adapts them to account for local structure and smoothness, [15] and [16] introduce the idea of a "hyper-GP" (using a stationary kernel) to predict the most probable length scale parameters to suit the local structure. These are used to make predictions on a non-stationary GP model of the terrain. [16] proposes to model space as an ensemble of GP's in order to reduce computational complexity.

While the context of this work is similar, the approaches differ. This work proposes the use of non-stationary kernels (neural network) to model large scale dis-continuous spatial data. It compares performances of stationary (squared exponential) and non-stationary (neural network) kernels for large scale data. It shows that using suitable non-stationary kernel can directly result in modeling local structure and smoothness, and hence suggests further research in this direction. It also proposes a local approximation methodology to emulate the locally adaptive effect of the techniques proposed in [14], [15] and [15]. This approximation technique is based on an efficient hierarchical representation (KD-Trees) of the data. Thus, the scalability issues relating to the application of this approach to large scale data sets are simultaneously addressed. The contribution of this work is thus the proposition of a novel way of representing large scale terrain using Gaussian Processes. Large scale field experiments are

used to support this claim. Such a representation naturally provides a multi-resolution model of the terrain, incorporates and handles uncertainty effectively and appropriately handles spatially dependent data.

III. Approach

A. Problem Definition

The terrain modeling problem can be understood as follows - given a sensor that provides terrain data as a set of points (x, y, z) in 3D space, the objectives of this work are to:

1) develop a multi-resolution representation that incorporates the sensor uncertainty in an appropriate manner.
2) be able to best handle the limitations of the sensor (eg. incomplete sensor information due to occlusions).

Terrain data can be obtained using numerous sensors including 3D laser scanners and GPS. 3D laser scanners provide dense and accurate data whereas a GPS based survey typically comprises of a relatively sparse set of well chosen points of interest. The experiments reported in this work use data sets obtained from both these sensors. The steps involved in creating the representation sought are depicted in Figure 1. The process of applying the GP representation to create elevation/surface maps at any arbitrary resolution is outlined in Figure 2.

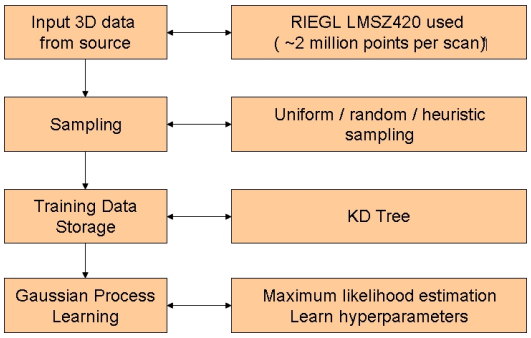

Fig. 1. The training / modeling process: A sensor such as a laser scanner is used to provide 3D terrain data. A sample of the given terrain data is used to learn a GP representation of the terrain. Different sampling methods can be used. Training data is stored in a KD-Tree for later use in the application stage. Learning the GP model of the terrain amounts to performing a maximum marginal likelihood estimation of the hyperparameters of the model. The outcomes of this process are the KD-Tree representation of the training data and the GP model of the terrain.

Typically, for dense and large data sets, not all of the data is required for learning the GP model. Such an approach would not scale due to computational complexity of the process. Thus, a sampling step may be included. This sampling may be of uniform nature, random and could also use heuristic approaches such as preferential sampling from areas of high gradient. The subset of the data that is to be used for training is stored in a KD-Tree [18] for later use in the inference process. The KD-Tree provides for rapid data access in the inference process and addresses scalability issues related to applying the proposed method in the context of large data sets.

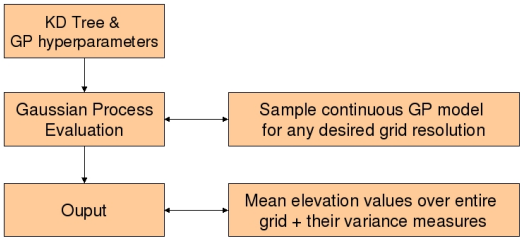

Fig. 2. The testing / application process: Given a desired region of interest where elevation needs to be estimated, applying the GP model learnt in the modeling stage amounts to sampling the continuous representation at the desired resolution, in the area of interest. A local approximation technique based on the KD-Tree representation obtained from the modeling stage ensures a good tradeoff between smoothness and feature preservation in the terrain. It also addresses scalability issues related to the method being applied for large scale data. The outcome of this process is an elevation / surface map.

B. Model Description

This work proposes the use of Gaussian Processes for modeling and representing terrain data. The steps below are directly based on [11]. Gaussian processes provide a powerful learning framework for learning models of spatially correlated and uncertain data. Gaussian Process Regression provides a robust means of estimation and interpolation of elevation information and can handle incomplete sensor data effectively. GP's are non-parametric approaches in that they do not specify an explicit functional model between the input and output. They may be thought of as a Gaussian Probability Distribution in function space and are characterized by a mean function $m(\mathbf{x})$ and the covariance function $k(\mathbf{x}, \mathbf{x}')$ where

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})], \tag{1}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \tag{2}$$

such that the GP is written as

$$f(\mathbf{x}) \sim \mathrm{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \tag{3}$$

The mean and covariance functions together specify a distribution over functions. In the context of the problem at hand, each $\mathbf{x} \equiv (x, y)$ and $f(\mathbf{x}) \equiv z$ of the given data.

The covariance function models the relationship between the random variables which, here, correspond to the given data. Although not necessary, the mean function $m(\mathbf{x})$ may be assumed to be zero by scaling the data appropriately such that it has an empirical mean of zero. There are numerous covariance functions (kernel) that can be used to model the spatial variation between the data points. The most popular kernel is the *squared-exponential* kernel given as

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^T \Sigma (\mathbf{x} - \mathbf{x}')\right), \tag{4}$$

where $k(\mathbf{x}, \mathbf{x}')$ is the covariance function or kernel; $\Sigma = \begin{bmatrix} l_x & 0 \\ 0 & l_y \end{bmatrix}^{-2}$ is the length-scale matrix, a measure of how

quickly the modeled function changes in the directions x and y; σ_f^2 is the signal variance. The set of parameters l_x , l_y , σ_f are referred to as the hyperparameters and specify what sort of values the parameters of the model might take.

Another kernel used in this work is the *neural network* kernel. It is specified by

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \arcsin\left(\frac{\beta + 2\mathbf{x}^T\Sigma\mathbf{x}'}{\sqrt{(1+\beta+2\mathbf{x}^T\Sigma\mathbf{x})(1+\beta+2\mathbf{x}'^T\Sigma\mathbf{x}')}}\right), \quad (5)$$

where β is a bias factor and l_x , l_y , σ_f , β are the hyperparameters used.

The main difference between these two kernels is that the squared-exponential kernel, being a function of $|\mathbf{x} - \mathbf{x}'|$ is stationary (invariant to translation) whereas the neural network function is not so. In practice, the squared exponential function has a smoothing or averaging effect on the data. The neural network covariance function proves to be much more effective than the squared exponential covariance function in handling dis-continuous (rapidly changing) data, as the experiments in this paper will show. This is the main reason why it proves to be very effective in modeling complex terrain data.

C. Learning the hyperparameters

Training the GP for a given data set is tantamount to optimizing the hyperparameters of the underlying kernel. For the squared-exponential kernel, this amounts to finding the optimal set of values for $\theta = \{l_x , l_y , \sigma_f , \sigma_n\}$ and for the neural network kernel, the optimal values for $\theta = \{l_x , l_y , \sigma_f , \beta , \sigma_n\}$, where σ_n^2 is the noise variance in the data being modeled. This is performed by formulating the problem in a Maximum Marginal Likelihood Estimation framework and subsequently solving a non-convex optimization problem.

Defining $X = \{\mathbf{x_i}\}_{i=1}^n = (x_i, y_i)_{i=1}^n$ and $\mathbf{z} = \{f(\mathbf{x}_i)\}_{i=1}^n = \{z_i\}_{i=1}^n$ as the sets of training inputs and outputs respectively with n instances, the log marginal likelihood of the training outputs ($\mathbf{z}$) given the set of locations (X) and the set of hyperparameters θ is given by

$$\log(\mathbf{z}|X,\theta) = -\frac{1}{2}\mathbf{z}^T K_z^{-1}\mathbf{z} - \frac{1}{2}\log|K_z| - \frac{n}{2}\log(2\pi), \quad (6)$$

where $K_z = K(X,X) + \sigma_n^2 I$ is the covariance matrix for all noisy targets $\mathbf{z}$ and $K(X,X)$ is covariance matrix for the noise-free targets (using equation 4 or 5). The log marginal likelihood has three terms - the first describes the data fit, the second term penalizes model complexity and the last term is simply a normalization coefficient. Thus, training the model will involve searching for the set of hyperparameters that enables the best data fit while avoiding overly complex models. Occam's razor [19] is thus in-built in the system and prevention of over-fitting is guaranteed by the very formulation of the learning mechanism. Using this maximum marginal likelihood formulation, training the GP model on a given set of data amounts to finding the optimal set of hyperparameters that maximize the log marginal likelihood. This can be performed using standard off-the-shelf optimization approaches. In this work, a combination of stochastic search (simulated annealing) and gradient descent (Quasi-Newton optimization with BFGS Hessian update [20]) was found to produce the best results. Using a gradient based optimization approach leads to advantages in that convergence is achieved much faster.

D. Applying the GP model

Applying the GP model amounts to using the learned GP model to estimate the elevation information across a region of interest, characterized by a grid of points at a desired resolution. The 2.5D elevation map can then be used as is or as a surface map for various applications. This is achieved by performing Gaussian Process Regression at the set of query points, given the training data set and the GP kernel with the learnt hyperparameters.

For additive independent identically distributed Gaussian noise with variance σ_n^2, the prior on the noisy observations becomes

$$\mathrm{cov}(z) = K(X,X) + \sigma_n^2 I. \quad (7)$$

The joint distribution of any finite number of random variables of a GP is Gaussian. Thus, the joint distribution of the training outputs f and test outputs f_* given this prior can be specified by

$$\begin{bmatrix} \mathbf{z} \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right). \quad (8)$$

For n training points and n_* test points, $K(X,X_*)$ denotes the $n \times n_*$ matrix of covariances evaluated at all pairs of training and test points and similarly for $K(X,X)$, $K(X_*,X_*)$ and $K(X_*,X)$. The function values (f_*) corresponding to the test locations (X_*) given the training inputs X, training outputs $\mathbf{z}$ and covariance function (kernel) is given by

$$\bar{f}_* = K(X_*,X)[K(X,X) + \sigma_n^2 I]^{-1}\mathbf{z} \quad (9)$$

and their uncertainty is given by

$$\mathrm{cov}(f_*) = K(X_*,X_*) - K(X_*,X)[K(X,X) + \sigma_n^2 I]^{-1}K(X,X_*) \quad (10)$$

Denoting $K(X,X)$ by K and $K(X,X_*)$ by K_*; for a single test point x_*, $k(x_*) = k_*$ the above equations can then be rewritten as:

$$\bar{f}_* = k_*^T(K + \sigma_n^2 I)^{-1}\mathbf{z} \quad (11)$$

and variance

$$V[f_*] = k(x_*,x_*) - k_*^T(K + \sigma_n^2 I)^{-1}k_* . \quad (12)$$

Equations 11 and 12 provide the basis for the elevation estimation process. The GP estimates obtained are a best linear unbiased estimate for the respective query points. Uncertainty is handled by incorporating the sensor noise model in the training data. The representation produced is multi-resolution in that a terrain model can be generated at

any desired resolution using the GP regression equations presented above. Thus, the terrain modeling approach proposed is a probabilistic, multi-resolution one that aptly handles spatially correlated information.

E. Fast Local Approximation using KD-Trees

As mentioned earlier, the squared exponential kernel has a smoothing effect on the data whereas the neural-network kernel is much more effective in modeling dis-continuous terrain data. In order to achieve a good trade-off between obtaining smooth terrain models and yet preserve the characteristic features in the terrain, a KD-Tree based local approximation methodology is proposed in this work. During the inference process, the KD-Tree that initially stored the training data is queried to provide a predefined number of spatially closest training data, to the point for which the elevation must be estimated. The GP regression process then uses only these training exemplars to estimate the elevation at the point of interest. The number of nearest neighbor exemplars used controls the tradeoff between smoothness and feature preservation and also the time taken for the inference process. Note that the GP model itself is learnt from the set of all training data but is applied locally using the KD-Tree approximation. Particularly, in the case of the neural network kernel, this amounts to adding the benefits of a stationary kernel to the highly adaptive power of a non-stationary kernel. This process provides two advantages - it tends to achieve the locally adaptive GP effect as exemplified by the works [14], [15] and [16] and it simultaneously addresses the scalability issue that arises when applying this approach to large scale data sets.

IV. Experiments

Experiments were performed on two data sets obtained using different sensory modalities. The first data set comprised of a scan from a RIEGL LMSZ420 laser scanner at the Mt. Tom Price mine in Western Australia. The data set comprised of about 1.85 million points in 3D space. The data set is depicted in Figure 3 and is referred to as the "Tom Price" data set in the rest of this paper. The second data set comprises of GPS based survey data (4612 points) from a large Kimberlite (diamond) mine. The second data set is a very hard data set in that it contains very sparse information (the average spread of the data is about 31m in x and y directions) spread over a very large geographical area ($\approx$ 2.2 x 2.3 sq km). The data set, depicted in figure 4, is referred to as the "Kimberlite Mine" data set in the rest of this paper.

Fig. 3. Laser Scanner data set taken at Mt. Tom Price mine in Western Australia using a RIEGL LMSZ420. The data set consists of over 1.85 million data points. The data set spanned about 135 x 72 sqm.

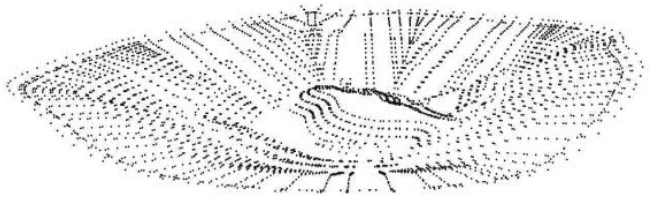

Fig. 4. GPS based survey data set from the Kimberlite (diamond) mine. The data set comprised of over 4600 points spanning about 2.2 x 2.3 sqkm. The data set is very sparse with the average spread being about 31m in x and y directions.

The focus of the experiments performed on these data sets was to understand the terrain modeling capabilities of Gaussian Processes. Specifically, the performance of stationary (squared-exponential) and non-stationary (neural network) kernel were compared. The Mean Squared Error (MSE) criterion was used in conjunction with visual inspection of the final output elevation/surface map to arrive at conclusions regarding the results obtained.

The results of testing various kernels on the two data sets are summarized in Tables I and II. Three different kernels were tested - stationary (squared exponential kernel denoted as SQEXP), non-stationary (neural network kernel denoted as NN2) and clamped neural network (denoted as NN2C), a variation wherein the noise parameter was bound to the known sensor uncertainty for the data set. The difference between the latter method and the first two is that whereas the first two methods incorporate the sensor uncertainty, the optimizer is allowed to use this parameter as an additional degree of freedom. In the latter most case however, the known uncertainty is incorporated and this parameter is fixed. The optimizer has 1 degree of freedom less than in previous cases. The table shows that Non-Stationary Kernel (Neural Network) outperforms Stationary Kernel (Squared Exponential) for both data sets. The clamped model performs comparably to the non-clamped one in the Tom Price data set and better, in the Kimberlite Mine data set. This is due to the fact that clamping the noise to the true sensor noise does not allow the optimizer to assign a more pessimistic value and thereby affect the output. The Neural Network GP is able to produce good terrain models even with one degree of freedom less (no noise hyperparameter as its fixed). The final output for the Tom Price data set is depicted in Figures 5 and 6; the corresponding confidence estimate is depicted in Figure 7. The output surface and confidence estimates for the Kimberlite Mine data set are respectively depicted in Figures 8 and 9 respectively.

TABLE I

Kernel performance: Tom Price dataset (1806944 points over 135 x 72 sq m, 3000 training data, 90198 test points for MSE)

Kernel	Mean Squared Error (MSE)
Squared Exponential (SQEXP)	0.085 sq m
Neural Network (NN2)	0.051 sq m
Constrained Neural Network (NN2C)	0.065 sq m

Numerous experiments were also performed to better understand the modeling process and the factors influencing

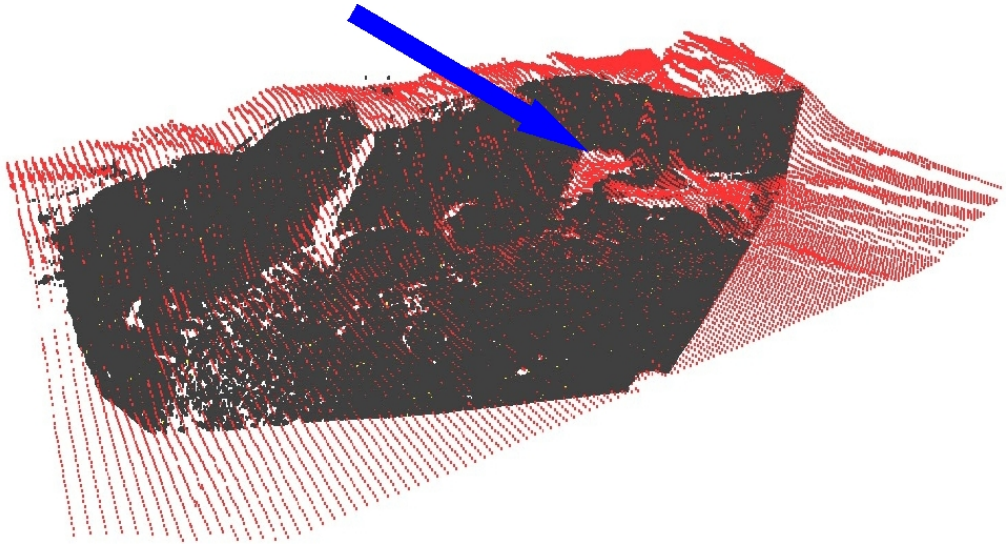

Fig. 5. Outcome of applying a single neural network based GP to the Tom Price data set. The gray (darker) points constitute the given data set. The red points are the output points - the elevation data as estimated by the GP over the test grid. The figure clearly demonstrates the ability of the GP to (1) reliably estimate the elevation data in known areas and (2) provide a good interpolated estimate in areas which have gaps

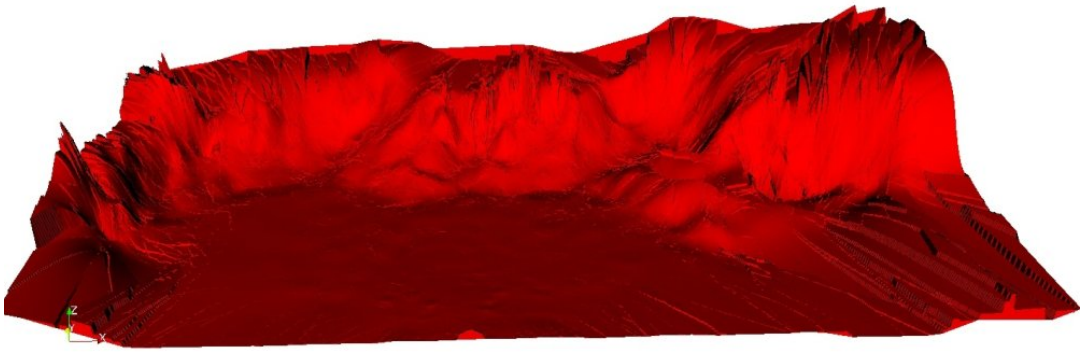

Fig. 6. The surface generated using the output elevation map obtained by applying a single neural network GP to the Tom Price data set. Refer Figure 5.

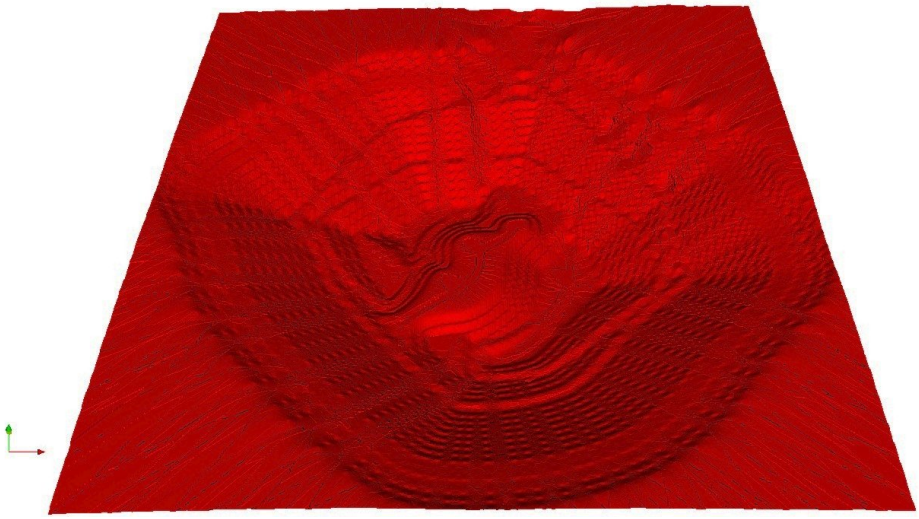

Fig. 8. Final Output Surface of the Kimberlite Mine data set obtained using a single Neural Network kernel based GP. The figure clearly demonstrates the ability of the GP to (1) reliably estimate the elevation data in known areas and (2) produce models that take into account the local structure so as to preserve the characteristics of the terrain being modeled.

it. These included issues like optimization strategy, sampling strategy, size of training data set and number of nearest neighbors for the fast local approximation step in the inference process. Details of these experiments are not provided here due to paucity of space - they are, however, available in an extended version of this paper which will be available online.

TABLE II

KERNEL PERFORMANCE: KIMBERLITE MINE DATASET (4612 POINTS SPREAD OVER 2.17 X 2.28 SQ KM)

Kernel	Number of training data	Mean Squared Error (MSE)
Squared Exponential (SQEXP)	1000	13.014 sq m over 3612 points
Neural Network (NN2)	1000	8.870 sq m over 3612 points
Squared Exponential (SQEXP)	4612	1.616 sq m over all 4612 points
Neural Network (NN2)	4612	1.191 sq m over all 4612 points
Constrained Neural Network (NN2C)	4612	0.880 sq m over all 4612 points

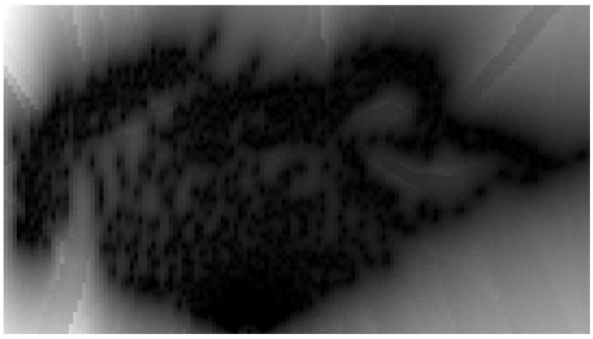

Fig. 7. Final Uncertainty Estimates of output data. Refer Figure 5. The areas in black are low uncertainty areas while the white areas have the highest uncertainty. The uncertainty is clearly more only in those areas where there was no observed / training data available - typically in the outer fringe areas and in gaps.

V. CONCLUSION

This work proposed the use of Gaussian Processes for modeling large scale and complex terrain. Specifically, it showed that a single neural network based Gaussian Process was powerful enough to be able to successfully model complex terrain data, taking into account the local structure and preserving much of the spatial features in the terrain. This work also proposed a fast local approximation technique based on KD-Trees that was able to simultaneously address the issues of balancing the tradeoff between smoothness and feature preservation as well as the issue of scalability when this approach is applied to large scale data sets. This work also compared stationary and non-stationary kernel and studied various detailed aspects of the modeling process. Experiments conducted on multiple real sensor data sets taken in a mining scenario clearly validated the claims made. The data sets included dense laser scanner data and sparse, feature rich GPS based survey data. The model obtained naturally provided a multi-resolution representation of large scale terrain, effectively handled both uncertainty and incompleteness in a statistically sound way and finally provided a powerful basis to handle correlated spatial data in an appropriate manner.

ACKNOWLEDGMENTS

This work has been supported by the Rio Tinto Centre for Mine Automation and the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

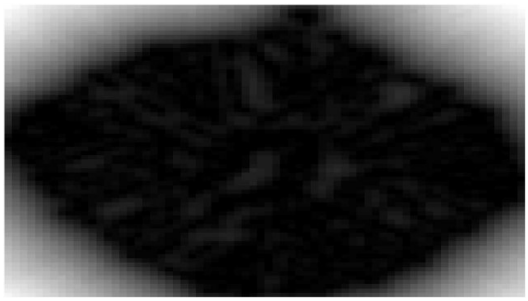

Fig. 9. Final Uncertainty Estimates of output data. Refer Figure 8. The areas in black are low uncertainty areas while the white areas have the highest uncertainty. The uncertainty is clearly more only in those areas where there was no observed / training data available - typically in the outer fringe areas and in gaps.

REFERENCES

[1] H. Durrant-Whyte, "A Critical Review of the State-of-the-Art in Autonomous Land Vehicle Systems and Technology," Sandia National Laboratories, USA, Tech. Rep. SAND2001-3685, November 2001.

[2] I. D. Moore, R. B. Grayson, and A. R. Ladson, "Digital terrain modelling: A review of hydrological, geomorphological, and biological applications," *Hydrological Processes*, vol. 5-1, pp. 3–30, 1991.

[3] E. Krotkov and R. Hoffman, "Terrain mapping for a walking planatary rover," *IEEE Transactions on Robotics and Automation*, vol. 10(6), pp. 728–739, 1994.

[4] C. Ye and J. Borenstein, "A new terrain mapping method for mobile robots obstacle negotiation," in *UGV Technology Conference at the SPIE AeroSEnse Symposium*, vol. 4, Orlando, FL, USA, April 2003.

[5] S. Lacroix, A. Mallet, D. Bonnafous, G. Bauzil, S. Fleury, M. Herrb, and R. Chatila, "Autonomous rover navigation on unknown terrains: Functions and Integration," *International Journal of Robotics Research (IJRR)*, vol. 21(10-11), pp. 917–942, 2002.

[6] R. Triebel, P. Pfaff, and W. Burgard, "Multi-Level Surface Maps for Outdoor Terrain Mapping and Loop Closing," in *International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, October 2006.

[7] J. Leal, S. Scheding, and G. Dissanayake, "3D Mapping: A Stochastic Approach," in *Australian Conference on Robotics and Automation*, November 2001.

[8] I. Rekleitis, J. Bedwani, D. Gingras, and E. Dupuis, "Experimental Results for Over-the-Horizon Planetary exploration using a LIDAR sensor," in *Eleventh International Symposium on Experimental Robotics*, July 2008.

[9] D. Kidner, M. Dorey, and D. Smith, "What's the point? Interpolation and extrapolation with a regular grid DEM," in *Fourth International Conference on GeoComputation*, Fredericksburg, VA, USA, 1999.

[10] I. S. Kweon and T. Kanade, "High-Resolution Terrain Map from Multiple Sensor Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14(2), February 1992.

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[12] P. K. Kitanidis, *Introdcution to Geostatistics: Applications in Hydrogeology*. Cambridge University Press, 1997.

[13] G. Matheron, "Principles of Geostatistics," *Economic Geology*, vol. 58, pp. 1246–1266, 1963.

[14] T. Lang, C. Plagemann, and W. Burgard, "Adaptive Non-Stationary Kernel Regression for Terrain Modeling," in *Robotics, Science and Systems*, 2007.

[15] C. Plagemann, K. Kersting, and W. Burgard, "Nonstationary Gaussian Process Regression using Point Estimates of Local Smoothness," in *European Conference on Machine Learning (ECML)*, 2008.

[16] C. Plagemann, S. Mischke, S. Prentice, K. Kersting, N. Roy, and W. Burgard, "Learning Predictive Terrain Models for Legged Robot Locomotion," in *International Conference on Intelligent Robots and Systems (IROS)*, 2008, to appear.

[17] C. J. Paciorek and M. J. Schervish, "Nonstationary Covariance Functions for Gaussian Process Regression," in *Neural Information Processing Systems (NIPS)*, 2004.

[18] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*. Springer, 1993.

[19] D. J. C. Mackay, *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, 2002.

[20] J. Nocedal and S. Wright, *Numerical Optimization*. Springer, 2006.