Non-Parametric Consistency Test for Multiple-Sensing-Modality Data Fusion

Marcos P. Gerardo-Castro*, Thierry Peynot†*, Fabio Ramos‡* and Robert Fitch*
*Australian Centre for Field Robotics (ACFR), The University of Sydney, NSW 2006, Australia.
†School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane QLD 4001, Australia.
‡School of Information Technologies, The University of Sydney, NSW 2006, Australia.
{m.castro,f.ramos,rfitch}@acfr.usyd.edu.au, t.peynot@qut.edu.au

***Abstract*—Fusing data from multiple sensing modalities, e.g. laser and radar, is a promising approach to achieve resilient perception in challenging environmental conditions. However, this may lead to *catastrophic fusion* in the presence of inconsistent data, i.e. when the sensors do not detect the same target due to distinct attenuation properties. It is often difficult to discriminate consistent from inconsistent data across sensing modalities using local spatial information alone. In this paper we present a novel consistency test based on the log marginal likelihood of a Gaussian process model that evaluates data from range sensors in a relative manner. A new data point is deemed to be consistent if the model statistically improves as a result of its fusion. This approach avoids the need for absolute spatial distance threshold parameters as required by previous work. We report results from object reconstruction with both synthetic and experimental data that demonstrate an improvement in reconstruction quality, particularly in cases where data points are inconsistent yet spatially proximal.**

I. Introduction

Advances in autonomous perception have enabled robots to operate outdoors in important applications such as agriculture, mining, defence, and autonomous driving. *Resilient* perception is necessary to support further advances in situations where robots must operate for long periods of time in challenging and variable environmental conditions. One way to achieve resilient perception is to employ multiple sensing modalities (MSMs) such as visual and infrared cameras [1], laser and camera [2], or laser and radar [3], [4].

As a system, the MSM approach can be resilient in cases that would severely compromise any single sensor acting alone. However, in some cases distinct sensing modalities can detect different targets even though they are spatially aligned. We then say that they provide *inconsistent* data, or *conflicting* data. This situation often leads to *catastrophic fusion* [5], where the quality of the representation of an object or scene obtained using traditional Bayesian data fusion is actually significantly degraded compared to the representations obtained using a single source of information [6].

Our recent work in fusing data from laser and radar established a *consistency test* to determine which subset of sensing data, across multiple sensing modalities, can be fused safely [6], [7]. This prevented the occurrence of catastrophic fusion. In this paper we improve on this prior work for data fusion in an MSM system by introducing an iterative consistency test that is entirely automatic (no hand-tuned parameter), and which significantly increases discriminatory power, particularly when data points from multiple sensors may appear to be locally consistent yet are inconsistent with respect to a global model.

The consistency test in our initial work [6], based on the Mahalanobis distance, provided encouraging results in avoiding catastrophic fusion in cases such as the presence of thick dust or smoke, which are often detected by lasers but not by radars. However, this work had some limitations and motivates further questions. Because the Mahalanobis distance essentially measures the "difference of opinion" between sensors within a local geometric neighbourhood, a threshold parameter is required in order to decide whether a pair of data points are consistent (and thus safe to incorporate) or not. We are interested in avoiding such a parameter. As a result, the need to choose parameter values for different situations will be removed and the consistency test will thus be more generally applicable.

The challenge in developing consistency tests arises due to differences between sensing modalities in terms of noise characteristics and resolution. It is thus important to maintain measures of uncertainty in the fused data. This case is in contrast to work that assumes homogeneity in these factors [8], [9], [10]. The specific challenge that we consider in this paper is how to perform consistency tests accurately while avoiding local threshold parameters.

In this paper we present a novel approach that uses an iterative consistency test based on the log-marginal likelihood of a Gaussian process (GP) model. We choose a single sensing modality as a reference, and evaluate whether data from other sensors statistically improve the reference model. This approach avoids local geometric threshold parameters and can be more discriminatory because it takes into account the global model and does not involve absolute distances; the comparison measures relative improvement of the GP model if the data point under consideration was to be fused. The assumption to bias the fusion towards one sensor that is trusted more than others is reasonable and can occur, for example, in a scenario with smoke or dust. Radar data are generally more immune from airborne contaminants than are laser data, but it is still beneficial to consider laser data where appropriate due to its

higher resolution.

The benefits of this approach can be seen in our results using both synthetic and experimental data. We evaluated our algorithm, in comparison to our previous method, in an object reconstruction task with two distinct range sensors acquiring 3D point clouds. The objects in the synthetic case have parts made of two different materials, one being transparent for only one of the sensors, and occluding the perception of the other sensor. Several scenarios with varying levels of difficulty are tested. The object in the experimental case is a car in an outdoor environment obscured by dust. Results show that the current method, with no hand-tuned parameter, performs at least as well as the previous method in all cases, and performs better in challenging cases where it is difficult to discriminate inconsistent data locally.

This paper is organised as follows. Related work is discussed in Sec. II. Section III describes the background data-fusion framework. Section IV then introduces the proposed method and Sec. V presents its experimental validation. Finally, Sec. VI concludes and discusses future work.

II. Related Work

The fusion of data acquired by multiple sensing modalities has been implemented in multiple domains. Examples include: fusing data from laser and radar [9], tactile and laser [11], [12], and ultrasonic and laser [13]. In this context it is typically assumed that physically aligned sensors detect the same target when pointed in the same direction [8], [9]. However, this assumption can lead to catastrophic fusion. Therefore, a consistency test is necessary to perform data fusion [6] when using distinct sensing modalities.

Approaches to consistency testing include blind source separation (BSS) methods [14] and dependency test methods [15]. With BSS methods, it is still necessary to resolve ambiguities and there are several restrictions on the mixing matrix structure. Dependency test methods must define the size or number of expected clusters, which is usually unknown. Although both methods can separate different sources from mixed measures, they do not natively build continuous representations or represent uncertainties with respect to the estimates, which are accounted for in our proposed method.

Robust data fusion approaches have been implemented in the context of Gaussian processes. A GP with a t-test prior has been used to avoid spurious data that affects the quality of continuous representations [16]. Our problem has been posed in a slightly different way. Rather than applying an outlier rejection to the data from a single sensing modality as is done in [16], we focus on analysing data from multiple sensing modalities by comparing models created with the GPs. In our problem, information from multiple modalities is given (i.e. we know the source of the data) and the goal is to detect when the data from two sensing modalities are consistent in order to perform robust data fusion.

Mixtures of Gaussian processes [17], [18] have also been proposed to compare heterogeneous models. However, the focus of application is different. In [17] and [18], the objective is to determine the trajectories of multiple targets, whereas in this work we want to estimate object or scene representations by fusing data from multiple sensing modalities.

In [6] we proposed a framework using Gaussian processes to estimate continuous surfaces with uncertainties from 3D data provided by distinct sensing modalities. The method compares two surfaces built from 3D points provided by each sensing modality individually, by applying a χ^2 test to the Mahalanobis distance between two distributions representing the two surfaces. The main drawback of this method is that a manually predefined threshold is required to determine when data are consistent or inconsistent, which leads to a lack of adaptability to new situations. In addition, the comparison of points from the two surfaces is achieved without any insight into the potential effect that fusing points from a distinct modality would have on the model. In this paper we propose an entirely automatic method, with no hand-tuned parameter, that takes data from one sensing modality and evaluates whether or not to fuse with points from a distinct sensing modality. This evaluation is based on the estimated impact that this addition will have on the model. If adding this point improves the model, it is considered consistent and fused. Otherwise it is declared inconsistent and set aside.

III. Background

A. Continuous Representations using Gaussian processes

In this work, we need to build continuous representations of objects or environments from data acquired by each sensing modality available. To that end we use GPs due to their ability to learn spatial correlations between noisy data in a non-parametric Bayesian fashion [19]. Assuming a single sensing modality i, the inputs of the GP are given by the vector $X^i = [\boldsymbol{x}_1^{i^T}, ..., \boldsymbol{x}_n^{i^T}]^T$, where $\boldsymbol{x}_k^{i^T} \in \mathbb{R}^d$, $\forall k$ is one input point, d is the dimension of each input point, and n is the number of training points. In this work d is usually equal to 1 (for range measurements in a plane), 2 (e.g. for an elevation map), or 3 for full $3D$ points, as in the experiments in Sec. V. The target data are given by $Y^i = [y_1^i, ..., y_n^i]^T$, where $y_k^i \in \mathbb{R}, \quad \forall k$. For example, in the classical $2D$ example of elevation maps, y_k would be the elevation at a position $\boldsymbol{x}_k$ in $2D$ space [20]. The GP provides a continuous representation of the output function f represented by the mean estimates $\bar{f}_*^{\,i}(\boldsymbol{x}_*)$ with uncertainties $\mathbb{V}^i(f_*(\boldsymbol{x}_*))$ which can be queried at any location $\boldsymbol{x}_* \in \mathbb{R}^d$. The predicted distribution is given by:

$$\mathbb{P}(f_*^i(\boldsymbol{x}_*) \mid X^i, Y^i, \theta^i, x_*) = \mathcal{N}(\bar{f}_*^{\,i}, \mathbb{V}^i[f_*]), \tag{1}$$

where θ^i are hyper-parameters. The mean $\bar{f}_*^i$ and variance $\mathbb{V}^i[f_*]$ at $\boldsymbol{x}_*$ given the measured data X^i are:

$$\bar{f}_*^i = k(\boldsymbol{x}_*, X^i)^T (K + {\sigma_i}^2 I)^{-1} Y^i \tag{2}$$

$$\mathbb{V}^i[f_*] = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{x}_*, X^i)^T (K + \sigma_i^2 I)^{-1} k(\boldsymbol{x}_*, X^i) \tag{3}$$

where K is a covariance matrix and σ_i^2 is the variance of the noise in the observed data. The hyper-parameters are $\theta^i = \{\Sigma_i, \ell_i, \sigma_i\}$, where ℓ_i is the length-scale of the data, and Σ_i^2 is the signal variance. Note that σ_i can be learnt along

with the other GP hyper-parameters. The optimisation of the hyper-parameters θ^i is done by maximising the log-marginal likelihood of the targets Y^i given the training inputs X^i. This log-marginal likelihood is given by:

$$LML(Y^i) = -\frac{1}{2}{Y^i}^T A^{-1} Y^i - \frac{1}{2} ln\ |A| - \frac{n}{2} ln\ 2\pi, \quad (4)$$

where we denote $LML(Y^i) \triangleq ln\ p(Y^i|X^i, \theta^i)$ and $A = K + {\sigma_i}^2 I$, for simplicity. The Cholesky decomposition is used to obtain the predictors ($\bar{f}_*$ and $\mathbb{V}[f_*]$) and the log-marginal likelihood [19].

In this work the input data are points acquired by range sensors and we focus on cases where each observation $\boldsymbol{x}$ is in 1D or 3D. In the 1D case we can use a Gaussian Beam Process (GBP) [21]. This method allows for a fully predictive model of range measurements $f_*^i(\boldsymbol{x}_*)$ to be built from only a few recorded range scans Y^i and bearing angles X^i. For the 3D case ($d = 3$) we use Gaussian Process Implicit Surfaces (GPIS) [22]. In this paper the experimental validation is focussed on the 3D case.

GPIS is a framework that models the surfaces of objects with complex geometry. In this framework, the input points X^i can be given as 3D point clouds, and the target values Y^i are the values of an implicit function f. In our implementation, the implicit surface of a 3D object is represented by a 0-level set function f defined such that $f(\boldsymbol{x})$ represents the signed distance between $\boldsymbol{x}$ and the surface of the object [23]. The values of $f(\boldsymbol{x})$ are positive for points inside the surface, and negative for points outside. The estimation of the surface is done by considering the observation of points on the surface ($f(\boldsymbol{x}) = 0$), which are usually the direct observations given by the range sensors, as well as some points inside and outside, also called constraints. GPIS follows the same formulation as described in Eq. (1), where the mean and variance are computed using Eqs. (2) and (3). 3D surface points and corresponding variances are then computed for zero values of $\bar{f}_*$ in Eq. (2) and Eq. (3) by querying points in a pre-defined region that covers the area of the observed object. The surface of an object of interest is reconstructed by computing the zero contour of $\bar{f}_*$.

B. Gaussian Process Data Fusion

For simplicity let us assume that we have observation data gathered by two sensing modalities i and j, denoted $(X^i, Y^i) \in \mathbb{R}^{n_i \times d} \times \mathbb{R}^{n_i}$ and $(X^j, Y^j) \in \mathbb{R}^{n_j \times d} \times \mathbb{R}^{n_j}$. In [9] we described a method to fuse *consistent* data from different sensing modalities using Gaussian Process Data Fusion (GPDF). The inputs of GPDF can be composed from the raw data of each sensing modality, e.g. $X = [X^i, X^j]$ and $Y = [Y^i, Y^j]$. Alternatively the inputs can be samples from continuous representations of data from each sensing modality e.g. $X = [S^i, S^j]$, where $S^k = \left\{\bar{f}^i(x_*), \mathbb{V}(f^i(x_*))\right\}$, and a set of target values $Y = [\boldsymbol{x}_*^i, \boldsymbol{x}_*^j]$. The variances are integrated to the GPDF as fixed noise parameters. The result of the fusion is a continuous representation with corresponding uncertainties

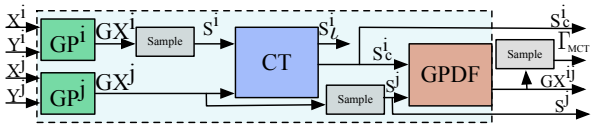


Fig. 1. The process of robust multiple-modality sensor data fusion presented in prior work [6]. This process takes data from sensing modality j, S^j, as a baseline and determines which subset S_c^j of data from modality i is *consistent* with S^j by applying a consistency test (CT) using the Mahalanobis distance. GPDF then fuses S^j with S_c^j. This results in a robust fused model GX^{ij} and a representation Γ_{MCT} that can be sampled as appropriate for any application.

expressed in $\bar{f}_*$ and $\mathbb{V}[f_*]$ respectively:

$$\bar{f}_* = {k_*}^T (K + G)^{-1} Y \quad (5)$$

$$\mathbb{V}[f_*] = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - {k_*}^T (K + G)^{-1} k_* \quad (6)$$

where G is a non-fixed noise matrix:

$$G = \begin{bmatrix} {\sigma_i}^2(X^i) I_{n_i} & 0 \\ 0 & {\sigma_j}^2(X^j) I_{n_j} \end{bmatrix} \quad (7)$$

As in [9] we integrate different noise models into the GP by implementing an input-dependent noise process following Eqs. (5) and (6). Both noise parameters can be specified, e.g. based on a predefined model, or they can be learnt along with the other hyper-parameters.

Note that in the original formulation of the method (in [9]) we assumed that the sensing modalities did not provide any *conflicting* data, i.e. that the two range sensors (a laser and a radar) always detected the same targets when pointing in the same direction. This was an appropriate assumption for the experiments conducted in [9]. However, in robotics there are many cases where this assumption is invalid, which led to the introduction of a consistency test in our GPDF framework.

C. GPDF with Consistency Test

In [6], we proposed a framework to perform robust data fusion that considers the difference of perception between different sensing modalities by introducing a consistency test applied on the sensor data in a GPDF framework. Consider that we start from data given by modality j. Given there may be conflicting data between the two sensing modalities, we need to determine which subset of data acquired by modality i should be fused with the data from modality j. The process is illustrated in Fig. 1.

First, the method uses a GPIS to generate a continuous representation model GX^i using data from sensing modality i only: $[X^i, Y^i]$. We then sample points S^i from the surface, which come in the form of points where the estimated mean $\bar{f}_* = 0$ (the estimated distance to the surface), with a corresponding variance. A continuous representation GX^j is also generated from the data of modality j, using another GPIS. The corresponding surface is S^j. A consistency test then evaluates if the sampled points S^i are consistent with the GX^j model. Points that pass this test are saved in the subset of consistent data S_c^i while points that fail the test are considered

as inconsistent data, saved in S_ι^i. We then use GPDF to fuse the data in S_c^i with samples from GX^j. This produces a final model GX^{ij} and a corresponding surface Γ_{MCT}.

In this prior work, the consistency test was formulated within a hypothesis testing framework. Consider the point currently tested to be $\boldsymbol{x}_*$. The hypothesis (H) was that $\boldsymbol{x}_*$ is located on S^i but not on S^j. To test our hypothesis we compared the two distributions expressed using Eq. (1) for modalities i and j, respectively. We used the Mahalanobis distance ($D(\boldsymbol{x}_*)$) to express the distance between the two distributions at $\boldsymbol{x}_*$. Considering that given H, D^2 has a χ^2 density with one degree of freedom, the validity of our hypothesis was subjected to a χ^2 acceptance test. Our hypothesis was thus tested using the following criteria:

$$H:\ D^2 > \chi^2_{1-t_\alpha}. \tag{8}$$

Therefore, we considered that $\boldsymbol{x}_*$ was on surface S^i but not on S^j if Eq. (8) was true. We used a significance level $t_\alpha = 0.05$, which gave us 95% probability concentration region of D^2.

In [6] we demonstrated the approach with data acquired by a laser (modality i) and a radar (modality j) in scenarios where the object (or scene) of interest is partly covered with a material that is detected by modality i but is transparent for modality j. An example of this scenario in field robotics is the presence of thick dust in the air, which is often detected by a laser but not by a radar, whose waves penetrate through. In this case, the inconsistent points S_ι^i correspond to the detection of dust by the radar. Using this proposed process, we were able to reconstruct a car surrounded by a significant amount of dust, by fusing radar data with consistent data from the laser.

IV. GPDF with Log-Marginal-Likelihood Consistency Test

The main drawback of the previous method is that a manually predefined threshold is required to determine how to separate the consistent data from the inconsistent data. Consequently, the method does not adapt well to different situations and scenarios, especially when the frontier separating the two types of data is very fine. In addition, all input points are tested *before* any fusion is executed, ignoring the actual impact that adding points from another modality will have on the fused model. In this paper, we propose to make the decision of whether to add points from the second sensing modality based on the likely impact it will have on the fused model. If adding a point leads to an improvement of the model (in terms of data fit) then the point is considered as consistent and is kept for fusion.

Our proposed method is based on an iterative evaluation of the log-marginal likelihood (LML) of the data. It incorporates a trade-off between model fit, model complexity and the number of points used. This trade-off is achieved by iteratively updating the model GX^j with sampled points from GX^i. Each sampled point from modality i is accepted in the model (i.e. accepted for fusion) if the LML of the model with the added point is greater than the LML without the point. The output of this process is a fused model (GX^{up}) that incorporates only the consistent data from modality i into the model GX^j.

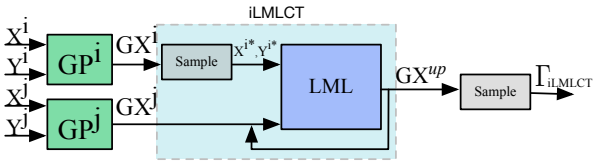


Fig. 2. GPDF with iLMLCT. k random samples taken from the model GX^i are tested for consistency with GX^j prior to fusion. If adding a sample improves the model then the point is fused and the model GX^{up} is updated accordingly. This improvement is tested using a log-marginal likelihood criteria. This process is repeated to test all k samples. The output of the process is the fused model GX^{up}, from which a continuous representation $\Gamma_{iLMLMCT}$ can be sampled as appropriate for any application.

Algorithm 1: Iterative LML Consistency Test Algorithm

Input: GX^i, GX^j, X^j, Y^j
Output: $GX^{up}, S_\iota^i, S_c^i$
Parameters: $\sigma_i, \sigma_j, k, K, L$
$[X^{i*}, Y^{i*}] = sample(GX^i, N_i)$
$[X^{up}, Y^{up}] = [X^j, Y^j]$
$GX^{up} = GX^j$
for $p = 1 : k$ **do**
  $Y^{j^+} = Y^{up} \cup Y^{i*}(p)$
  $X^{j^+} = X^{up} \cup X^{i*}(p)$
  $K^* = KUpdate(K, X^{j^+}(p))$
  if $LML(Y^{j^+}) > LML(Y^{up})$ **then**
    $[L, GX^{up}] = GXUpdate(GX^{up}, K^*, \sigma_j, \sigma_i, L)$
    $Y^{up} = Y^{j^+}$
    $X^{up} = X^{j^+}$
    $S_c^i = S_c^i \cup [X^{i*}(p), Y^{i*}(p)]$
    $K = K^*$
  else
    $S_\iota^i = S_\iota^i \cup [X^{i*}(p), Y^{i*}(p)]$

A. Iterative Consistency Test using LML

The proposed process of iterative log-marginal likelihood consistency (iLMLCT) test is illustrated in Fig. 2 and Algorithm 1. Similarly to the process described above in Sec. III-C, we start with a GPIS model built from data from modality j only, and consider which part of the data from modality i should be integrated in a fused model. First a model GX^i is built from the modality i data only, which provides a continuous representation, and also the sensor data noise characteristics. N_i samples ($[X^{i*}, Y^{i*}]$) are then randomly taken from GX^i (on the surface at $f = 0$ in the 3D case). In practice, in this paper we used $N_i = 2\,n_i$, where n_i is the number of input points from modality i, i.e. the size of X^i. Then, for each sample, one by one, we evaluate the potential impact the addition of this sample would have on the model. This is tested by comparing the log-marginal likelihood of

the current model $LML(Y^{up})$ (from Eq. (4)) with the log-marginal likelihood after adding the new point, $LML(Y^{j^+})$:

$$LML(Y^{j^+}) = -\frac{1}{2}{Y^{j^+}}^T {B^*}^{-1} Y^{j^+} - \frac{1}{2} ln\,|B^*| - \frac{n+1}{2} ln\,2\pi, \tag{9}$$

where $Y^{j^+} = Y^{up} \cup Y^{i*}(p)$ is the new set of targets, and $B^* = K^* + diag(\sigma_j^2, ..., \sigma_j^2, \sigma_i^2)$. K^* is the updated covariance matrix, and is obtained by augmenting K with the covariance values between the new point $X^{i*}(p) \in \mathbb{R}^d$ and the set X^j (see in Line 7 of Algorithm 1).

If $LML(Y^{j^+}) > LML(Y^{up})$ then the new point is considered as consistent, therefore, it is added to the subset S_c^i and will be fused in the model. Thus, the set of points X^{up}, Y^{up} and model GX^{up} including the Cholesky factor L and covariance K, will be updated by adding the new observation. Updating the model (see details below in Section IV-B) is done by using the function $GXUpdate$ which approximates the inversion of an updated K^* by computing a Cholesky factor L'.

On the other hand, if the LML with the added point is lower, then the point is considered inconsistent, therefore, it is added to S_ι^i and is not labelled for fusion in the model. The process is repeated for each of the N_i samples taken from modality i. Once the process is completed, we have an improved continuous model GX^{up} that is the result of the fusion of data from modality j with the full subset of consistent data from modality j. The model can then be sampled as needed by the application, for example to generate a surface Γ^{up} in the 3D case.

B. Updating the Model with New Points

Since this method requires updating our GP model iteratively, and building a new GP model is computationally expensive, it is important to perform this operation in an efficient way. We use the Cholesky decomposition to reduce the computational cost from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ [24] when incorporating a new point ($\boldsymbol{x}_*$) into a covariance matrix K. This is calculated as shown in Eq. (10) and (11), where $K_{3,3} = k(\boldsymbol{x}_*, \boldsymbol{x}_*)$ and c_3 is the solution of the linear system $L_{c3} = k_{1,2}$, where $c_3 = [a\ b]^T$:

$$M = \begin{bmatrix} K_{1,1} & K_{1,2} \\ K_{1,2} & K_{2,2} \end{bmatrix}, L = \begin{bmatrix} C_{1,1} & C_{1,2} \\ 0 & C_{2,2} \end{bmatrix} \tag{10}$$

$$L' = \begin{bmatrix} C_{1,1} & C_{1,2} & a \\ 0 & C_{2,2} & b \\ 0 & 0 & chol(K_{3,3} - c_3^T c^3) \end{bmatrix} \tag{11}$$

The mean estimates $\bar{f}_*$ and variances $\mathbb{V}[f_*]$ are computed using the new Cholesky factor L:

$$\bar{f}_* = K_*^T \alpha, \tag{12}$$

$$\mathbb{V}[f_*] = \frac{L'}{K_*}, \tag{13}$$

where $\alpha = L^T \backslash (L' \backslash Y^{up})$. This process is repeated for each of the N_i sampled points from GX^i that need to be added to the model.

V. Experimental Validation

We evaluated the ability of the approach to generate accurate reconstructions of the surfaces of objects in the environment by fusing data acquired by two distinct sensing modalities. Both sensors have been spatially aligned a priori, and in all experiments the targets detected by the sensors are static in the form of 3D point clouds. We first evaluated the proposed approach using synthetic data of objects scanned by virtual sensors (Sec. V-B), then using real experimental data of objects scanned by sensors on-board an outdoor mobile robot (Sec. V-C). In this section, we first describe the metrics used to analyse the results, then we describe and analyse the results obtained using simulated and experimental data.

A. Performance Metric

To evaluate the performance of the proposed robust Data Fusion approach, in the experimental results below we compute the accuracy of the fused continuous representations obtained after robust fusion and compare them with the method in prior work and with the reconstructions achieved with only one type of sensing modality. To calculate the error between a set of samples from an estimated surface S_* and a ground-truth (GT) surface (e.g a CAD model), S_{GT}, we compute the root-mean-squared error (RMSE) of the Euclidean distance between each point $X_*(p)$ on S_* and the closest point on S_{GT}. Naming this distance $dist(X_*(p))$ and considering N sample points gives:

$$RMSE = \sum_{p=1}^{N} \frac{\sqrt{dist^2(X_*(p))}}{N}. \tag{14}$$

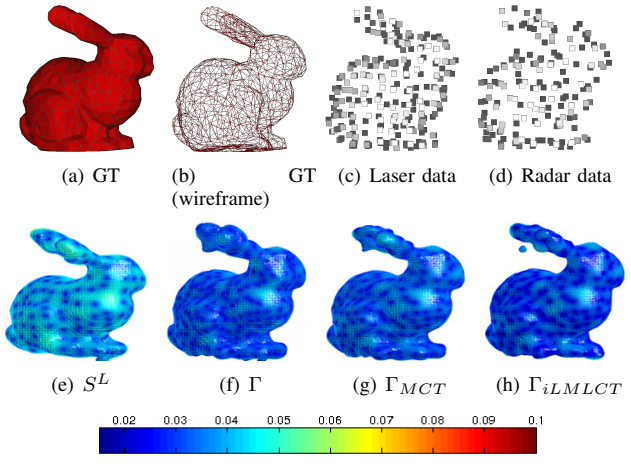
In the experiments below we use $N = 10,000$ samples.

B. Simulated Objects Test case

We first validated the ability of the approach to accurately reconstruct the surface of 3D objects ($d = 3$) by using synthetic data and setting up different challenging scenarios, in particular due to the presence of a material that is transparent for one sensor but not for the other. We show that the method can accurately reconstruct the surface of the objects even when they are partially covered by this transparent material.

1) Testing Conditions: Synthetic objects were scanned by two virtual sensors namely $i = L$ (virtual laser) and $j = R$ (virtual radar), with $X^L \in \mathbb{R}^{n_L \times d}$ and $L \in \mathbb{X}^{\mathbb{R}^{n_R \times d}}$. We considered objects composed of two different materials, φ_1 and φ_2. In this context, sensor L is able to reliably detect both φ_1 and φ_2. However, the material φ_1 is transparent for R, which is only able to detect φ_2. For these simulations, CAD models of the objects made of φ_2 were used as ground truth.

2) Space-Bunny Results: In the first test scenario we use data from the Stanford-bunny [25] in different conditions. To analyse the difference of perception during data fusion we tested two different scenarios where the Stanford-bunny is equipped with different sizes of a helmet that is partially covering the top of the bunny. We assume that the Stanford-bunny is made of φ_2, with a spherical helmet made of φ_1, i.e. transparent for modality R (virtual radar). The considered



(a) GT (b) GT (wireframe) (c) Laser data (d) Radar data

(e) S^L (f) Γ (g) Γ_{MCT} (h) Γ_{iLMLCT}

(i) Uncertainty: 2 std. deviations ($2\sqrt{\mathbb{V}[f_*]}$), in m

Fig. 3. *Stanford Bunny* Results. Both sensing modalities consistently observe the same object. (a) Ground truth: the Stanford Bunny, made of material φ_2 (in red). (b) Wireframe representation. 3D point clouds from the simulated scans are shown in (c) for sensor L ($n_L = 330$ points) and (d) for sensor R ($n_R = 166$ points). (e-h) show reconstructions using $GPIS$, coloured by uncertainties, from blue to red, where red is the highest level of uncertainty (see (i)). (e) $GPIS$ reconstruction using only data from sensor L. (f) $GPIS$ reconstruction using GPISDF without consistency test. (g) $GPISDF$ reconstruction with consistency test using Mahalanobis Distance (MCT). (h) $GPISDF$ reconstruction after consistency test with LML.

TABLE I
RMSE OF GPIS OF STANFORD-BUNNY (IN M)

	RMSE $\pm$ std. dev.
S^L	0.017 $\pm$ 0.014
S^R	0.049 $\pm$ 0.034
Γ	**0.016 $\pm$ 0.012**
Γ_{MCT}	0.017 $\pm$ 0.013
Γ_{iLML}	0.017 $\pm$ 0.013
	% inconsistent points
MCT	3%
iLMLCT	2%

noise level was taken as $\sigma_L = 0.03m$ for sensor L and as $\sigma_R = 0.1m$ for sensor R. We call this object the *Space Bunny*. The bunny fits in a box of dimensions $1.4 \times 1.5 \times 0.6\, m^3$.

To get some initial insight into the performance of the the consistency test methods we first tested a control case where the bunny has no helmet, hence there is no significant difference of perception between the sensors (see Fig. 3). Because both sensing modalities are able to detect materials φ_2, the method should find that all points from both sensing modalities are consistent. Any point found to be inconsistent would represent a false alarm. Table I quantifies the results obtained for the Stanford-Bunny in terms of RMSE and standard deviation (std. dev.) of the error over the N samples, as defined in Sec. V-A. The GT is the full-resolution surface representation of the bunny. Results for laser reconstruction S^L, radar reconstruction S^R, fusion without consistency test Γ, fusion with Mahalanobis distance consistency test (MCT) Γ_{MCT} and fusion with LML test Γ_{iLML}, are compared.

TABLE II
RMSE OF GPIS OF SPACE-BUNNY WITH BIG HELMET (IN M)

	RMSE $\pm$ std. dev.
S^L	0.102 $\pm$ 0.152
S^R	0.049 $\pm$ 0.036
Γ	0.165 $\pm$ 0.164
Γ_{MCT}	0.017 $\pm$ 0.024
Γ_{iLML}	**0.017 $\pm$ 0.020**
	% inconsistent points
MCT	38%
iLMLCT	39%

TABLE III
RMSE OF GPIS OF SPACE-BUNNY WITH SMALL HELMET (IN M)

	RMSE $\pm$ std. dev.
S^L	0.074 $\pm$ 0.079
S^R	0.049 $\pm$ 0.036
Γ	0.073 $\pm$ 0.080
Γ_{MCT}	0.030 $\pm$ 0.019
Γ_{iLML}	**0.029 $\pm$ 0.019**
	% inconsistent points
MCT	23%
iLMLCT	27%

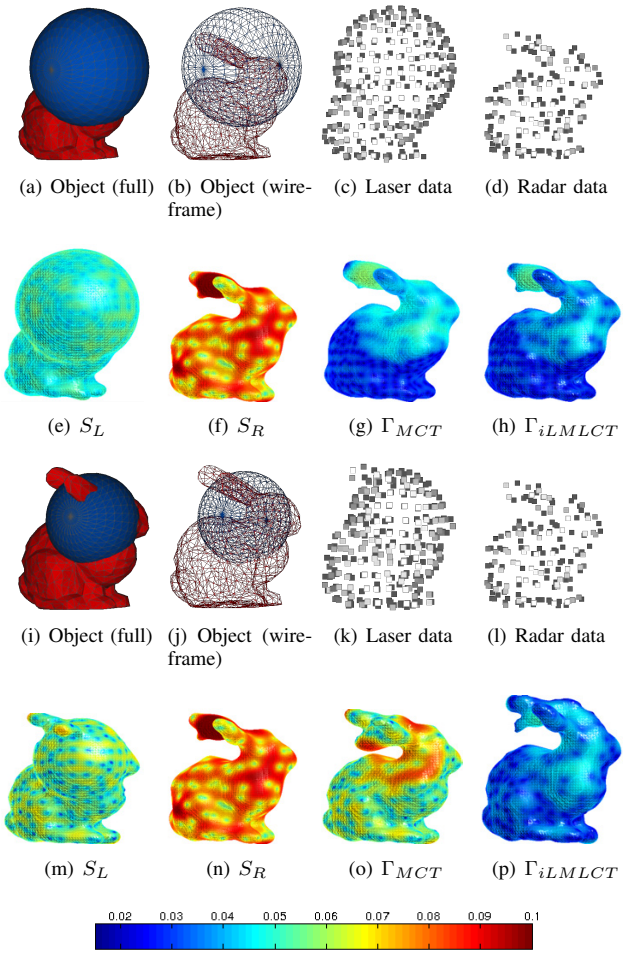
We can see that the accuracy obtained by the fusion methods with consistency test is comparable with the regular fusion without the test, which confirms that almost all points were found consistent. The table also shows that, as expected, only a very small fraction of L points (2%) were found inconsistent and excluded from the fusion points by the iLML test.

Fig. 4 shows the Space Bunny. We considered two scenarios, by varying the dimensions of the helmet. The first case shown in Fig. 4 is the *Space-Bunny-Big-Helmet* (a-h), where the helmet covers the whole top of the bunny. This object fits in a box of dimensions $1.5 \times 1.8 \times 0.8m^3$. The second scenario is *Space-Bunny-Small-Helmet* (i-p). It is even more challenging, with a smaller helmet that does not cover the ears completely, and also intersects with the head of the bunny. The *Space-Bunny-Small-Helmet* fits in a box of dimensions $1.4 \times 1.5 \times 0.7m^3$.

Table II shows the RMSE results for the Space-Bunny-Big-Helmet. The objective is to accurately reconstruct the bunny itself, despite the presence of the helmet. Therefore, the ground-truth surface in this case was that of the bunny alone. We can see that the fusion methods with consistency test are again nearly three times as accurate as what could be achieved with the virtual radar data alone ($RMSE = 0.017m$ vs. $0.049m$), and 10 times more accurate than the fusion that integrates all points from both sensing modalities (i.e. without consistency test). Note that the results obtained with the MCT and the new iLML method are comparable, but the uncertainty is slightly lower with the latter. To reach those results, both methods excluded a significant percentage of the L points from fusion (close to 40%), as they were found to be inconsistent.

Table III shows the RMSE results for the Space-Bunny-Small-Helmet. Again both fusion methods with consistency tests produce an accurate representation of the surface of the bunny, but in this more challenging situation, the new iLML method performs better, with slightly lower RMSE, and this was obtained without hand-tuning any threshold. To reach those results, the new method excluded more inconsistent points from fusion than the previous method (27% vs. 23%).

3) Knot-Oval Results: In the second test scenario we consider the task of reconstructing the surface of a knot made of φ_2 despite the presence of a polygonal ellipsoid made of φ_1, i.e. transparent for sensor L, and with noisier data:



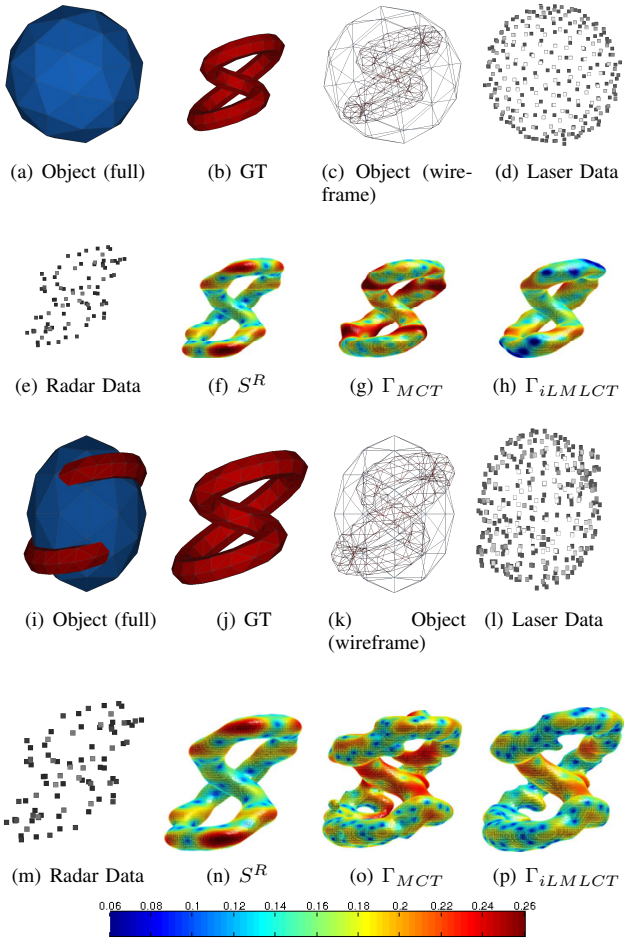
(a) Object (full) (b) Object (wireframe) (c) Laser data (d) Radar data

(e) S_L (f) S_R (g) Γ_{MCT} (h) Γ_{iLMLCT}

(i) Object (full) (j) Object (wireframe) (k) Laser data (l) Radar data

(m) S_L (n) S_R (o) Γ_{MCT} (p) Γ_{iLMLCT}

Fig. 4. Variants of the *Space-Bunny* object. (a) and (i): The synthetic object composed of a spherical helmet made of material φ_1 (in blue) on top of the Stanford bunny, made of material φ_2 (in red, GT). (b) and (j): Wireframe representation, showing the part of the Bunny occluded by the helmet. 3D point clouds from the simulated scans are shown in (c) (482 points) and in (k) (251 points) for sensor L, and in (d) and (l) for sensor R (166 points). (e-h) and (m-p) show reconstructions using $GPIS$, coloured by uncertainties, from blue to red, where red is the highest level of uncertainty. (e) and (m): $GPIS$ reconstruction using only data from sensor L. (f) and (n): $GPIS$ reconstruction using only data from sensor R. (g) and (o): $GPISDF$ reconstruction with MCT. (h) and (p): $GPISDF$ reconstruction after consistency test with LML.



(a) Object (full) (b) GT (c) Object (wireframe) (d) Laser Data

(e) Radar Data (f) S^R (g) Γ_{MCT} (h) Γ_{iLMLCT}

(i) Object (full) (j) GT (k) Object (wireframe) (l) Laser Data

(m) Radar Data (n) S^R (o) Γ_{MCT} (p) Γ_{iLMLCT}

Fig. 5. *Knot-Oval* Results. (a) and (i): Full objects, including a polygonal ellipsoid made of material φ_1 (in blue). (b) and (j): The object to reconstruct: knot made of material φ_2 (in red, GT). (c) and (k): Wireframe representation of the full object. 3D point clouds are shown in: (d) ($n_L = 282$ points) and (k) for ($n_L = 343$ points) for L data, and (e) and (m) for R data ($n_R = 85$ points in both cases). (f-h) and (n-p): Reconstructions using $GPIS$, coloured by uncertainties, from blue to red, as shown in the colour bar. (f) and (n): $GPIS$ reconstruction using only data from Sensor R. (g) and (o): $GPISDF$ reconstruction using MCT. (h) and (p): $GPISDF$ reconstruction using iLML.

$\sigma_L = 0.06m$ and $\sigma_R = 0.25m$. In the first case considered, the ellipsoid covers the knot entirely, see Fig. 5(a-e). As a result, modality R only perceives the knot while sensor L only perceives the ellipsoid around it. This means that effectively all sensor L data points are inconsistent with sensor R data points. We name this object *Knot-Big-Oval*. The object fits in a box of dimensions $2.99 \times 3.36 \times 2.35\, m^3$. Table IV reports the quantified results for the surface reconstructions shown in Fig. 5(f-h). The RMSEs show that the reconstruction of the knot by the method using the MCT test is worse than the surface reconstructed with radar data only (S^R), while the surface obtained with the new test iLML is almost as accurate as S^R. This is most likely due to the fact the MCT method only rejected 92% vs. 97% for iLML. In some places the ellipsoid's surface is very close to the knot's, making the consistency test challenging, especially for MCT considering the high level of noise in the R data.

TABLE IV
RMSE OF GPIS OF KNOT-BIG-OVAL (IN M)

	RMSE ± std. dev.
S^R	0.055 ± 0.051
Γ_{MCT}	0.069 ± 0.065
Γ_{iLML}	**0.058 ± 0.054**
	% inconsistent points
MCT	92%
iLMLCT	97%

TABLE V
RMSE OF GPIS OF KNOT-SMALL-OVAL (IN M)

	RMSE ± std. dev.
S^R	0.055 ± 0.051
Γ_{MCT}	0.089 ± 0.086
Γ_{iLML}	**0.037 ± 0.039**
	% inconsistent points
MCT	78%
iLMLCT	76%

The second case is even more challenging: the polygonal

Fig. 6. The UGV equipped with laser and radar sensors used in this work.

Fig. 7. The UGV (left) observing the car (right) surrounded by dust.

ellipsoid is slightly smaller and its surface intersects with the knot in many places, see Fig. 5(i-l). Modality R still perceives the knot only, while sensor L perceives the ellipsoid and a few sections of the knot (Fig. 5(i)). We name this object *Knot-Small-Oval*. Table V quantifies the accuracy of the surface reconstructions shown in Fig. 5(n-p). Once again the reconstruction of the knot when using the MCT test is worse than S^R. On the other hand, the RMSEs obtained indicate that the surface obtained with iLML is significantly more accurate than S^R. Although the shape of S^R looks reasonably accurate in Fig. 5(n), the knot is actually too thin compared with the ground truth and Γ_{iLML}.

C. Real-World Experiments

1) Experimental Setup: We also tested the proposed approaches using real experimental data extracted from the datasets in [26], which were collected using an unmanned ground vehicle (UGV) (see Fig. 6) equipped with two range scanners (laser and mm-wave radar) and a cm-accuracy 6-DOF dGPS/INS localisation unit. The laser was a 2D Sick LMS291, with a 180° field of view (FOV), 0.25° angular resolution and a range resolution of $0.01m$. The mm-wave radar was a $94GHz$ Frequency Modulated Continuous Wave (FMCW) radar. Its field of view (FOV) is 360°, but it was restricted to the front view in these experiments, to be comparable with the laser. The radar's angular resolution is 2° and range resolution is $0.2m$. The two sensors were roughly aligned at a fixed tilt angle, and then calibrated to determine the actual transformation between them. To acquire the $3D$ data, the platform was driven around a rural environment, scanning objects multiple times from multiple perspectives from distances varying from $2m$ up to $30m$. Dust was introduced into the scene, affecting the perception of the laser scanner, which consistently detected airborne dust particles.

2) Data Preparation: Laser and radar data were pre-processed as described in [9]. The result is a set of 3D points per scan, similar to the data provided by a multi-echo laser sensor. Laser and radar raw scans were then cropped to only keep data where the two sensors' FOVs overlap. Laser and radar points were then transformed into a common global navigation frame. This transformation was obtained by combining the output of a prior extrinsic sensor calibration (using the technique in [27]) with the localisation of the UGV. The object of interest was then manually segmented from the full point cloud obtained with each sensing modality.

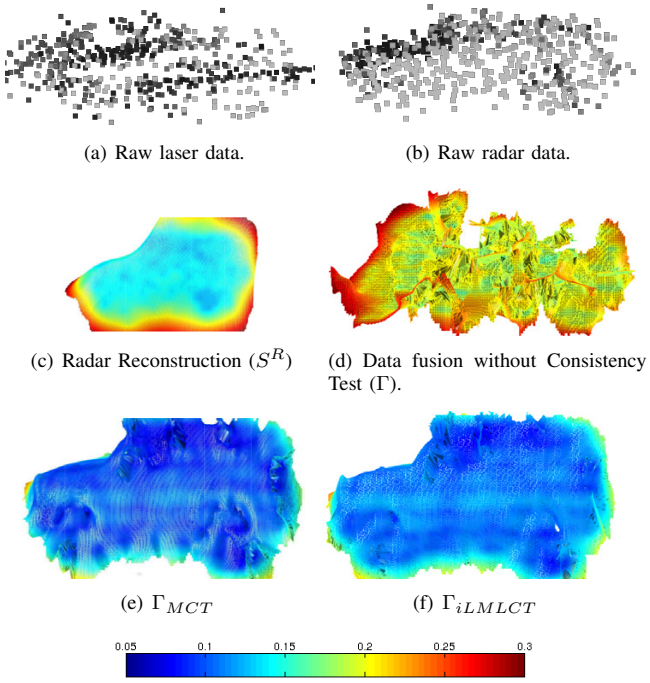
3) Experimental Results: We followed the proposed process to perform a robust $GPIS$ data fusion, where individual laser and radar surfaces (S^L and S^R, respectively) were first generated and then subjected to the proposed consistency test. Since we operated in environments with airborne dust, in these experiments we considered the radar as the baseline sensing modality (i.e. modality j in the algorithm described in Sec. IV) and we used the consistency test to determine which points from the laser should be fused into the model.

A car covered with airborne dust was scanned by our UGV (see Fig. 7). The car's surface was then reconstructed using $GPIS$ for each sensing modality, and using our proposed method. Figs. 8(a) and 8(b) show the raw data acquired by the laser and radar, respectively. Fig. 8(d) shows that GPISDF without consistency testing generates an unrecognisable shape, and has high uncertainty levels. This is because many dust points were fused together with points from the car. On the other hand, the proposed GPIS robust data fusion (see Fig. 8(f)) was able to recover the basic shape of the car, without the inconsistent data from the dust, and also to dramatically reduce the uncertainty levels of the estimates.

VI. Conclusions and Future Work

In this paper, we proposed a new data fusion method based on Gaussian Processes that is appropriate for data provided by distinct sensing modalities. The method introduces a non-parametric data consistency test based on the iterative evaluation of the log-marginal likelihood of the data. We provided experimental analysis comparing the results of 3D surface reconstructions with other algorithms including single-modality reconstructions and our previous method based on the Mahalanobis distance. The proposed method showed a substantial improvement in the accuracy of surface representations and a reduction of uncertainty, especially in challenging cases when conflicting data between different sensing modalities were spatially close to each other.

In future work, we consider building highly accurate ground-truth of objects scanned by UGVs in the field to further evaluate the impact of this method in field robotics scenarii. We will also extend the experimental analysis by using other performance metrics such as the negative log probability



(a) Raw laser data. (b) Raw radar data.

(c) Radar Reconstruction (S^R) (d) Data fusion without Consistency Test (Γ).

(e) Γ_{MCT} (f) Γ_{iLMLCT}

Fig. 8. Experimental results. Estimation of the surface of a car covered with dust, as shown in Fig. 7. Surface reconstructions are coloured by uncertainties, from blue to red, where red is the highest level of uncertainty. (a) Raw laser data. (b) Raw radar data. (c) $GPIS$ Radar Reconstruction surface estimate. The reconstructed car fits in a box of dimensions $3.25 \times 1.89 \times 5.02\, m^3$. (d) $GPISDF$ surface estimate without consistency test, showing parts of the car and the dust cloud, The reconstructed car fits in a box of dimensions $4.49 \times 2.32 \times 6.93\, m^3$. (e) Surface reconstruction obtained with the GPISDF with the MCT method. (f) Surface reconstruction obtained with the GPISDF with the proposed iLMLCT method.

(NLP), which takes prediction variance into account [19]. One of the limitations of the proposed method is that covariance matrix updates will be more computationally expensive as more points are added to the reference model. We could reduce the number of initial points to be tested by performing a preliminary consistency test such as suggested in prior work [6], with a very tolerant criteria. Another improvement can be obtained by using sparse approximations [28]. Finally, although the results in this paper were obtained using two sensing modalities, the framework could also be used for situations with a larger number of sensing modalities.

REFERENCES

[1] C. Brunner, T. Peynot, and T. Vidal-Calleja, "Combining multiple sensor modalities for a localisation robust to smoke," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, San Francisco, CA, September 2011.

[2] T. Peynot and A. Kassir, "Laser-camera data discrepancies and reliable perception in outdoor robotics," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Taipei, Taiwan, October 2010.

[3] A. Kelly, A. Stentz, O. Amidi, M. Bode, D. Bradley, A. Diaz-Calderon, M. Happold, H. Herman, R. Mandelbaum, and T. Pilarski, "Toward reliable off road autonomous vehicles operating in challenging environments," *Int. J. of Robot. Res.*, 2006.

[4] T. Peynot, J. Underwood, and S. Scheding, "Towards reliable perception for unmanned ground vehicles in challenging conditions," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2009.

[5] J. Movellan and P. Mineiro, "Robust sensor fusion: Analysis and application to audio visual speech recognition," *Mach. Learn.*, 1998.

[6] M. P. Gerardo-Castro, T. Peynot, F. Ramos, and R. Fitch, "Robust Multiple-Sensing-Modality Data Fusion using Gaussian process implicit surfaces," in *IEEE Int. Conf. on Information Fusion*, 2014.

[7] M. P. Gerardo-Castro and T. Peynot, "Laser-to-radar sensing redundancy for resilient perception in adverse environmental conditions," in *ARAA Australasian Conf. on Robotics and Automation*, 2012.

[8] S. Dragiev, M. Toussaint, and M. Gienger, "Uncertainty aware grasping and tactile exploration," in *IEEE Int. Conf. on Robotics and Automation*, 2013.

[9] M. P. Gerardo-Castro, T. Peynot, and F. Ramos, "Laser-radar data fusion with Gaussian process implicit surfaces," in *Int. Conf. on Field and Service Robotics*, 2013.

[10] S. Vasudevan, "Data fusion with Gaussian processes," *Robot. Auton. Syst.*, 2012.

[11] J. Ilonen, J. Bohg, and V. Kyrki, "Fusing visual and tactile sensing for 3-d object reconstruction while grasping," in *IEEE Int. Conf. on Robotics and Automation*, 2013.

[12] J. Mahler, S. Patil, B. Kehoe, J. van den Berg, M. Ciocarlie, P. Abbeel, and K. Goldberg, "GP-GPIS-OPT: Grasp planning under shape uncertainty using Gaussian process implicit surfaces and sequential convex programming," in *IEEE Int. Conf. on Robotics and Automation*, 2015.

[13] T. Vidal-Calleja, D. Su, F. De Bruijn, and J. Miro, "Learning spatial correlations for bayesian fusion in pipe thickness mapping," in *IEEE Int. Conf. on Robotics and Automation*, 2014.

[14] C. Zhang and I. Sato, "Image-based separation of reflective and fluorescent components using illumination variant and invariant color," *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 2013.

[15] L. Song, A. Smola, A. Gretton, and K. M. Borgwardt, "A dependence maximization view of clustering," in *Int. Conf. on Machine learning*, 2007.

[16] P. Jylänki, J. Vanhatalo, and A. Vehtari, "Robust Gaussian process regression with a student-t likelihood," *J. Mach. Lean. Res.*, 2011.

[17] V. Tresp, "Mixtures of Gaussian processes," in *Neural Information Processing System*, 2000.

[18] M. Lázaro-Gredilla, S. Van Vaerenbergh, and N. Lawrence, "Overlapping mixtures of Gaussian processes for the data association problem," *Pattern Recognition*, 2012.

[19] C. Rasmussen and C. Williams, *Gaussian processes for Machine Learning*. The MIT Press, 2006.

[20] S. Vasudevan, F. Ramos, E. Nettleton, and H. Durrant-Whyte, "Gaussian process modeling of large-scale terrain," *J. Field Robot.*, 2009.

[21] C. Plagemann, K. Kersting, P. Pfaff, and W. Burgard, "Gaussian beam processes: A nonparametric bayesian measurement model for range finders," in *Robotics: Science and Systems III*, 2007.

[22] O. Williams and A. Fitzgibbon, "Gaussian process implicit surfaces," in *Gaussian processes in practice Workshop*, 2007.

[23] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans, "Reconstruction and representation of 3D objects with radial basis functions," in *Conf. on Computer Graphics and Interactive Techniques*, 2001.

[24] V. Guizilini and F. Ramos, "Online self-supervised segmentation of dynamic objects," in *IEEE Int. Conf. on Robotics and Automation*, 2013.

[25] G. Turk, "The Stanford bunny," http://www.gvu.gatech.edu/people/faculty/greg.turk/bunny/bunny.html, 2000.

[26] T. Peynot, S. Scheding, and S. Terho, "The Marulan data sets: Multi-Sensor perception in natural environment with challenging conditions," *Int. J. of Robot. Res.*, 2010.

[27] J. P. Underwood, A. Hill, T. Peynot, and S. J. Scheding, "Error modeling and calibration of exteroceptive sensors for accurate mapping applications," *J. Field Robot.*, 2010.

[28] E. Snelson and Z. Ghahramani, "Local and global sparse Gaussian process approximations," in *Int. Conf. on Artificial Intelligence and Statistics*, 2007.