



This is a repository copy of *A Gaussian Process Regression Approach for Fusion of Remote Sensing Images for Oil Spill Segmentation*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/131616/>

Version: Accepted Version

Proceedings Paper:

Longman, F.S., Mihaylova, L. and Yang, L. (2018) A Gaussian Process Regression Approach for Fusion of Remote Sensing Images for Oil Spill Segmentation. In: Proceedings of the 21st IEEE International Conference on Information Fusion. 21st International Conference on Information Fusion, 10-13 Jul 2018, Cambridge, UK. IEEE . ISBN 978-0-9964527-6-2

<https://doi.org/10.23919/ICIF.2018.8455304>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

A Gaussian Process Regression Approach for Fusion of Remote Sensing Images for Oil Spill Segmentation

Fodio S Longman, Lyudmila Mihaylova and [†] Le Yang

Department of Automatic Control and System Engineering,
University of Sheffield, Mappin Street, S1 3JD, UK

Email: fslongman1@sheffield.ac.uk*, l.s.mihaylova@sheffield.ac.uk, le.yang.le@gmail.com[†]

[†]Department of Electrical and Computer Engineering,
University of Canterbury, Christchurch, New Zealand.

Abstract—Synthetic Aperture Radar (SAR) satellite systems are very efficient in oil spill monitoring due to their capability to operate under all weather conditions. This paper presents a framework using Gaussian process (GP) to fuse SAR images of different modalities and to segment dark areas (assumed oil spill) for oil spill detection. A new covariance function; a product of an intrinsically sparse kernel and a Rational Quadratic Kernel (RQK) is used to model the prior of the estimated image allowing information to be transferred. The accuracy performance evaluation demonstrates that the proposed framework has 37% less RMSE per pixel and a compelling enhancement visually when compared with existing methods.

Index Terms: Oil Spill, Synthetic Aperture Radar (SAR), Registration, Image Fusion, Segmentation, Gaussian Processes

I. INTRODUCTION

Oil spills are caused by accidental discharge or illegal dumping of oil ballasts by oil vessels and drilling platforms into the environment, causing enormous damage both socially and environmentally [1]. Marine oil spill for example, pollutes sea water, destroy wildlife, coastal beaches and affects the overall quality of life of marine inhabitants, raising concerns on oil transportation across the sea and a growing interest in developing efficient methods for oil spill detection [2].

The Macando blow-out that occurred on the 20th April 2010 is of national significance in the United States. It was an accidental oil spill caused by the explosion and sinking of the Deep Water Horizon offshore platform making the sea floor oil gusher to flow for 87 days and releasing more than 200 million gallons of oil into the Gulf of Mexico, resulting in loss of lives and damage to the marine ecosystem. The company responsible, British Petroleum (BP) was made to pay the largest environmental fine in history, a total of about 18.7 billion dollars [3]. Early detection of oil spill will help towards efficient disaster management. To detect and monitor oil spill, remote sensing systems with sensors on-board a satellite or aircraft are used to acquire images of the earth from distance. Sensors in different bands of the electromagnetic spectrum have been applied, e.g. in hyperspectral and multi spectral bands, or Synthetic Aperture Radar (SAR) for oil spill

detection in marine environment [1], [4]–[7]. However, SAR images are the most widely used for this purpose as they are not affected by local weather conditions and cloudiness [8].

SAR is an active microwave sensor that acquires two dimensional (2D) images [9]. The performance of detecting oil spill in a SAR image, largely depends on sea conditions and the ability of oil films to decrease the backscattering of the sea surface, resulting in dark formations. A general assumption is that oil spill appears as dark areas on SAR images due to the dampen effects on capillary waves [10], [11]. However, not all dark formations are oil spills, necessitating the need for a robust detection technique and verification. Fig 1. shows an example of two challenging dark formations for detection:

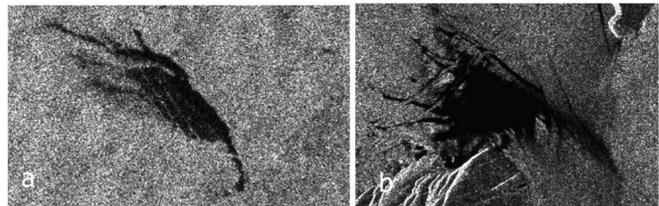


Fig. 1: Example of challenging dark formations: (a) verified oil spill acquired 6/09/2005, Ancona Italy, (b) Verified look-alike acquired 25/08/2005, Otranto, Italy. [10]

Recent developments have enabled more spaceborne SAR systems (e.g. the Sentinels) to be designed and launched, providing the scientific community with wide range of data in multi-modal configuration, including multi-frequency (C, L, X etc.), multi-angle ($10^\circ - 70^\circ$), multi-polarisation(dual, quad, etc) and multi-resolution features. With the availability of multi-modal SAR data, new methods to detect oil spill can be explored by combining data originating from different sources, with the aim of obtaining information of greater quality than individual sensor data used in previous studies for oil spill detection [5], [8]–[10], [12]–[15].

Fusion of SAR images, however, imposes several challenges due to multi-modalities, differences in sensor characteristics

and image acquisition modes. The individual images need to be registered in space and time. A suitable fusion algorithm needs to be chosen that will efficiently bring together the complimentary information from the individual modalities [7]. Additionally, it is important to define the level of fusion since data fusion can be performed at different information levels including pixel, feature and decision level [16].

Previous studies have reported significant improvement in oil spill classification, segmentation and discrimination with fused SAR images compared to using the individual images alone. In [17], fusion of SAR and hyperspectral images (HSI) is performed at pixel level, although the fusion method used is not described in the paper, the approach focused on fusing images from different sensors (SAR and HSI) and not on multi-frequency SAR data. The works of [7], [18] explored fusion of multi-frequency (S&X-bands) and multi-resolution (C-band) SAR images by adopting the wavelet transform approach.

Wavelet transform improves the spatial resolution of the fused image while preserving the colour appearance for interpretation [19], this is important since oil spill appears as dark formation on SAR images. With wavelets, images are converted from the spatial domain to the frequency domain and then decomposed into approximation and detail coefficients while preserving information, allowing image properties to be transferred using a fusion rule. However, if the decomposition scale is small, mosaic effects occurs on the fused image. On the other hand, if the scale is large, the colour contents of the fused images are lost, making it unsuitable for images with different spectral channels [20].

This paper proposes to solve oil spill segmentation problems by fusing multi-resolution SAR images using a Gaussian process regression approach. The approach is based on the design of a non-stationary covariance kernel to handle the change of support problem that exists in multi resolution images, The approach extends the work from [21] over different image modalities. A prior covariance function, the product of an intrinsically sparse covariance kernel and a rational quadratic kernel is utilised to model the high resolution pixel coordinates and their intensity values, forming a base covariance from which the new modality image is constructed. We consider that Gaussian process models have been used in object recognition in situations where the images are in different resolutions with the training data [22]. Additionally, GP priors are adaptable for inter-modality data encoding with multiple output behaviour [21]. The aim is to construct an image with high spatial and high spectral resolution.

The rest of the paper is organised as follows. In section II, we present the proposed framework and a brief description of the registration is provided, with more detailed explanation of the process already discussed in our previous work [7]. Subsequently, GP is introduced and the proposed kernel design and fusion process are described. In addition to this, we describe the performance validation measures of the approach. The K-means segmentation is described in III and lastly, we discuss the results in IV and present concluding remarks in V.

II. PROPOSED FRAMEWORK

The proposed framework is shown in Fig 2. It comprises a pre-processing step which performs image filtering, for reducing speckles inherent in SAR images, and image enhancement, for improving visualisation to obtain the best possible image perception, respectively. The next process is an automatic image registration that aligns the images so that a common spatial frame is realised. Lastly, the fusion and performance evaluation stages complete the system framework.

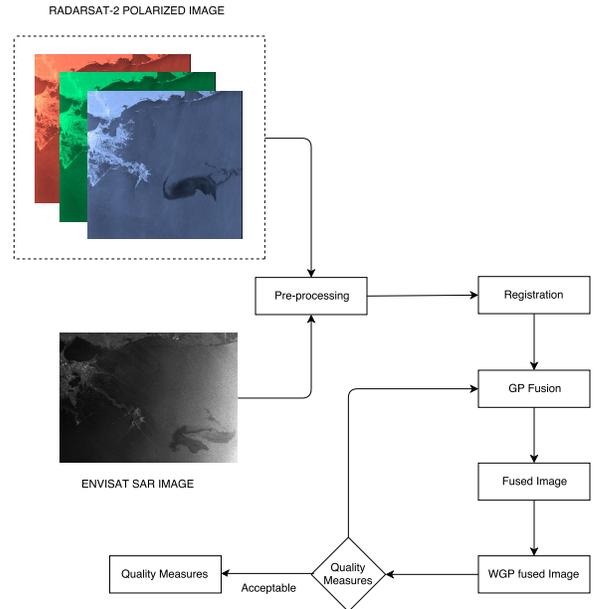


Fig. 2: GP Fusion Framework

A. Registration

The purpose of image registration is to establish correspondence between the images to be fused and to determine the geometric transformation that aligns one image with the other [23]. In this paper, the registration is done using Scale Invariant Feature Transform (SIFT) algorithm [24]. SIFT is a feature based registration that allows extraction and matching of distinct features from images. This is achieved in the following steps:

1) *Scale Space Extrema Detection*: Is the product of convolution between the variable scale Gaussian kernel $G(x, y, \sigma)$ and the image $\tilde{I}(x, y)$, given as

$$L(x, y, \sigma) = G(x, y, \sigma) * \tilde{I}(x, y) \quad (1)$$

Here, $*$ is the convolution operator with respect to pixel coordinates x and y of the image and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

is the Gaussian filter. To detect features in the image [25] proposed the use of scale space extrema in the Difference-of-Gaussian (DoG) convolved with the image, such that

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * \tilde{I}(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3)$$

where k is the multiplicative factor usually set to $\sqrt{2}$ [26]. Local maxima and minima of DoG are found by comparing each sample point to its eight neighbours in the image and nine neighbours in the scale and below.

2) *Key point Localisation*: This step involves fitting found key points to a nearby data for location, scale and rotation. The purpose is to remove noise-sensitive points or non-edge points and enhance stability of the matching process to improve immunity to noise.

3) *Orientation Assignment*: In this step, the location information is extracted from key points with identified position and scale, this is achieved by computing the pixel differences using the magnitude of the gradient $m(x, y)$ and the orientation $\theta(x, y)$.

4) *Key point Descriptor*: This phase computes the descriptors for the image region making it highly distinctive and invariant to illumination and angle changes. This is achieved by placing a Gaussian window over the regions and adding each sample to form orientation histograms, summarising the contents over an 8×8 sub region as shown in Fig.6 and 7 in [7].

5) *Feature Matching*: The purpose of this step is to find correspondence between the detected feature points. This is achieved by identifying nearest neighbours in the database of key points created from the extraction step. In the presence of outliers, Random Sample Consensus (RANSAC) is used. RANSAC is a robust technique that handles transformation estimation [25]. It is given by

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = s \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (4)$$

where (x_1, y_1) and (x_2, y_2) are the control points coordinates in the images to be registered. The variables t_x and t_y are translational values in x and y direction, s is the scaling factor and θ is the angle of rotation.

B. Gaussian Processes

Gaussian Processes (GP) are determined by a mean function and a covariance function also known as the covariance kernel. The mean $m(\mathbf{x})$ and the covariance $k(\mathbf{x}, \mathbf{x}')$ of a space function $f(\mathbf{x})$ are given as

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \quad (5)$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \quad (6)$$

and the GP can be described as:

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (7)$$

GP is a stochastic process, defined as a collection of random variables [27]. For convenience, the mean function is often assumed a zero value since GP can be adjusted to model the mean swiftly [28], while the covariance kernel is determined by some hyperparameters. A detailed explanation on kernels and hyperparameter adaptation is discussed in [27]. To achieve the mapping of inputs to an output space, GP imposes a

Gaussian prior distribution over the space functions $f(\mathbf{x})$, to map inputs $\mathbf{x}_i \in \mathbb{R}^D$ to the output space $\mathbf{y}_i \in \mathbb{R}$, where the output \mathbf{y} is a noisy observation represented as

$$\mathbf{y} = f(\mathbf{x}) + \epsilon \quad (8)$$

where

$$\epsilon \sim \mathcal{N}(0, \sigma^2) \quad (9)$$

is a Gaussian distribution with a zero mean and standard deviation σ .

To make predictions, GP learns the hyperparameters from the given training dataset here represented as \mathbf{Q} . The training is done on N input-output pairs such that: $\mathbf{Q} = \{(\mathbf{y}_1, \mathbf{x}_1), \dots, (\mathbf{y}_N, \mathbf{x}_N)\}$, where the function values are normally distributed with the modelled mean and covariance defined as

$$[f(\mathbf{x}_1)^T f(\mathbf{x}_2)^T \dots f(\mathbf{x}_N)^T]^T = \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (10)$$

here, $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$ are as defined in (5) and (6), respectively.

Representing the number of observations N as $\{\mathbf{X}, \mathbf{y}\}$, where $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^D\}_{i=1}^N$ and $\mathbf{y} = \{y_i \in \mathbb{R}\}_{i=1}^N$ and test points M given as $\{\mathbf{X}^*\} = \{\mathbf{x}_i^* \in \mathbb{R}^D\}_{i=1}^M$, the joint density of the observations N and the test points M is given as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \mathcal{N}\left(\boldsymbol{\mu} \begin{pmatrix} \mathbf{X} \\ \mathbf{X}^* \end{pmatrix}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & K(\mathbf{X}, \mathbf{X}^*) \\ K(\mathbf{X}^*, \mathbf{X}) & K(\mathbf{X}^*, \mathbf{X}^*) \end{bmatrix}\right) \quad (11)$$

Here, $\boldsymbol{\mu}(\mathbf{x})$ is the mean function and $k(\mathbf{x}, \mathbf{x}')$ is a positive semi-definite covariance function. From (11) above, the predictive distribution of the mean and covariance functions can be defined as

$$p(\mathbf{y}^* | \mathbf{X}, \mathbf{y}, \mathbf{X}^*) \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*) \quad (12)$$

where

$$\boldsymbol{\mu}_* = k(\mathbf{X}^*, \mathbf{X}) K_X^{-1} \mathbf{y} \quad (13)$$

$$\boldsymbol{\Sigma}_* = k(\mathbf{X}^*, \mathbf{X}^*) - k(\mathbf{X}^*, \mathbf{X}) K_X^{-1} k(\mathbf{X}, \mathbf{X}^*) \quad (14)$$

Here, \mathbf{K}_X is defined as:

$$\mathbf{K}_X = k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} \quad (15)$$

and $\sigma_n^2 \mathbf{I}$ is the sensor measurement noise, \mathbf{I} is an N -dimensional identity matrix. Subsequently, we will revert to our earlier notations of the mean ($\boldsymbol{\mu}_*$) as $m(\mathbf{x})$ and the covariance ($\boldsymbol{\Sigma}_*$) as $k(\mathbf{x}, \mathbf{x}')$. Considering that the $m(\mathbf{x})$ of the process is not always confined to a zero value, the mean function can be modelled to conveniently express the prior information allowing the predictive mean to be written as

$$\mathbf{y}^* = \mathbf{m}(\mathbf{X}_*) + K(\mathbf{X}_*, \mathbf{X}) K_X^{-1} (\mathbf{y} - \mathbf{m}(\mathbf{X})), \quad (16)$$

where $\mathbf{K}_X = K + \sigma^2 \mathbf{I}$ and $K = k(\mathbf{X}, \mathbf{X})$.

1) *Covariance Function*: The covariance function or kernel plays the central role in a GP. It encodes the inference of the underlying process by defining the correlation between function values [29]. In this paper, we aim to construct an image with a new modality, by combining SAR images with different image resolutions. To achieve this, firstly we require the covariance function to handle the change of support problem that exists in multi-modal images by conducting inference over image pixels with different resolutions. This is achieved by extending the observation point kernel to adapt to a multiple task kernel function over spaces, and utilising an integral kernel derivation described in [30]. The assumption here is there exist a 2D GP function \mathbf{f} within the spectral channels of the low resolution image, such that the designed model of a pixel is the result of observing the output function \mathbf{f}_* over areas of the high resolution image $k(\mathbf{H}_A, \mathbf{H}'_A)$ rather than points $k(\mathbf{x}, \mathbf{x}')$ which is the norm in a standard GP. A simple average relationship is then established between observed pixels and \mathbf{f}_* . A detailed derivation of defining covariance over areas is described in [21].

Secondly, a prior of the new modality image structure is defined. The covariance function prior design is based on the spatial characteristic of the high resolution image, forming the base for the new modality image construction. For the fusion problem, Let \mathbf{H}_A denote the locations of the high-resolution image pixels and \mathbf{L}_A the locations of the low-resolution image pixels, the covariance between two high-resolutions image pixels is defined as

$$k(\mathbf{H}_A, \mathbf{H}'_A) = \frac{1}{|\mathbf{H}_A||\mathbf{H}'_A|} \iint_{\mathbf{x} \in \mathbf{H}_A} \iint_{\mathbf{x}' \in \mathbf{H}'_A} k(\mathbf{x}, \mathbf{x}') dx dx' \quad (17)$$

where $k(\mathbf{H}_A, \mathbf{H}'_A)$ defines the covariance between two high resolution image areas and $|\mathbf{H}_A|$ is the surface area of \mathbf{H}_A

To design the prior, we consider that image data are normally non-smooth, and exhibit discontinuities with spatial non-stationarity. For this reason, the spatial information of the high-resolution image is used as the input space of the covariance function, added with the observed pixels of the high-resolution image to achieve contextual non-stationarity and to address image discontinuity problem. In the first step of the prior design, we exploit an intrinsically sparse covariance function proposed by [31] to obtain a sparse covariance. The sparse kernel is smooth but not infinitely differentiable making it suitable for application exhibiting discontinuities. A detailed description and derivation of this kernel is given in [31]. Let V_S be an intrinsically sparse kernel defined as

$$V_S(\mathbf{H}_A, \mathbf{H}'_A; \sigma_0, l) = \begin{cases} \sigma_0 \left[\frac{2 + \cos(2\pi \frac{d}{l})}{3} (1 - \frac{d}{l}) + \frac{1}{2\pi} \sin(2\pi \frac{d}{l}) \right] & \text{if } d < l \\ 0 & \text{if } d \geq l \end{cases} \quad (18)$$

where d is the distance between the midpoint coordinates of

the high-resolution pixel areas here defined as

$$d = |\text{mid}(\mathbf{H}_A) - \text{mid}(\mathbf{H}'_A)| \quad (19)$$

The variables σ_0 and l are the parameters of the kernel. where σ_0 determines the average distance of the function from the mean, l is the characteristic length scale that determines the length of change of the function.

In the second step, we link the observed high-resolution image pixels using a Rational Quadratic Kernel (RQK). The RQK is equivalent to adding together several Squared Exponential (SE) with different length-scales, enabling smoothness transfer and efficiency in handling the change of support problem. Let V_P be the RQK defined as

$$V_P(\mathbf{I}_H(\mathbf{H}_A), \mathbf{I}_H(\mathbf{H}'_A); \alpha, \theta_p) = \sigma^2 \left(1 + \frac{(\mathbf{I}_H(\mathbf{H}_A) - \mathbf{I}_H(\mathbf{H}'_A))}{2\alpha\theta_p^2} \right)^{-\alpha} \quad (20)$$

Here $\mathbf{I}_H(\mathbf{H}_A)$ and $\mathbf{I}_H(\mathbf{H}'_A)$ represent the pixel intensity values in \mathbf{H}_A and \mathbf{H}'_A , respectively. The variables α and θ_p are the parameters of the function. The function V_S provides a smooth, sparse and neighbouring covariance kernel, while V_P link image pixels within the covariance based on related information contained in the high-resolution image. Additionally, the RQK function V_P enables image smoothness transfer and handles the change of support problem. Hence, the image prior covariance function is then defined as a product of two independent processes (a) A positive semi-definite spatial covariance V_S and (b) A positive semi-definite pixel intensity covariance function V_P

$$k(\mathbf{H}_A, \mathbf{H}'_A) = \sigma^2 V_S((\mathbf{H}_A, \mathbf{H}'_A; \theta) V_P(\mathbf{I}_H(\mathbf{H}_A), \mathbf{I}_H(\mathbf{H}'_A); \theta) \quad (21)$$

Using the high-resolution areas ($\mathbf{H}_A, \mathbf{H}'_A$) and the augmented intensity values ($\mathbf{I}_H(\mathbf{H}_A), \mathbf{I}_H(\mathbf{H}'_A)$) of the high-resolution image pixels, the GP model is then trained to learn the hyperparameters represented as θ of the image prior $k(\mathbf{H}_A, \mathbf{H}'_A)$ as described in the next section.

2) *Hyperparameter Adaptation*: Hyperparameters refer to the parameters of the prior $k(\mathbf{H}_A, \mathbf{H}'_A)$ that includes parameters of the mean, covariance and noise term $\sigma_n^2 \mathbf{I}$. Here the parameters of the model are defined as $\theta = \{\theta_s, \theta_p, \sigma_f, \alpha, l\}$, where σ_f is an amplitude hyperparameter while θ_p controls sensitivity, α determines the relative weighing for large-scale and smaller-scale variations. When $\alpha \rightarrow \infty$ the behaviour of the kernel is identical to the SE kernel. The parameters of the prior need to be selected appropriately as they determine the quality of the output image. To optimize the hyperparameters, the Bayes approach is considered because it allows the use of continuous optimization methods enabling faster computation [28]. To achieve this, the marginal likelihood is maximised such that

$$p(\mathbf{y}|X) = \int p(\mathbf{y}|\mathbf{f}, X) p(\mathbf{f}|X) d\mathbf{f} \quad (22)$$

From (12), the likelihood $\mathbf{y}|\mathbf{f} \sim \mathcal{N}(\mathbf{f}, \sigma_n^2 \mathbf{I})$, with the GP prior over the latent function f from (11), gives the log of the marginal likelihood

$$\log(p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})) = -\frac{1}{2}\mathbf{y}^T K_X^{-1} \mathbf{y} - \frac{1}{2} \log |K_X| - \frac{n}{2} \log 2\pi. \quad (23)$$

In (23), the first term of the log of marginal likelihood finds data fit, the second is a model complexity term while the third is a constant, making it robust to over-fitting.

The optimized parameters of the covariance function from (23) are used to calculate $k(\mathbf{H}_A, \mathbf{H}'_A)$ which forms the base of the new modality image we aim to construct. Using the prior covariance, we find the covariance $k(\mathbf{H}_A, \mathbf{L}'_A)$ that couple the high-resolution pixels with the observed low-resolution pixels. First, the observed \mathbf{L}_A pixels are approximated by the \mathbf{H}_A pixels, an integration over \mathbf{L}_A with respect to point \mathbf{x} , the sum of the piecewise integration over the pixels of \mathbf{H}_A [21], defined as

$$k(\mathbf{H}_A, \mathbf{L}'_A) = \frac{1}{T_H} \sum_{\mathbf{H}'_A \in \mathbf{L}'_A} k(\mathbf{H}_A, \mathbf{H}'_A) \quad (24)$$

where T_H is the number of high-resolution \mathbf{H}_A areas that are contained in \mathbf{L}_A area. Accordingly, the corresponding covariance between the low-resolution pixels is defined as

$$k(\mathbf{L}_A, \mathbf{L}'_A) = \frac{1}{T_H T'_H} \sum_{\mathbf{H}_A \in \mathbf{L}_A} \sum_{\mathbf{H}'_A \in \mathbf{L}'_A} k(\mathbf{H}_A, \mathbf{H}'_A) \quad (25)$$

3) *Image Fusion*: To fuse the two image modalities, the training data of the model comes from the low-resolution image comprising the \mathbf{L}_A spatial areas and the pixel intensity values $\mathbf{I}_L(\mathbf{L}_A)$ that were observed. The GP model is then queried over the high resolution \mathbf{H}_A spatial areas where the $\mathbf{I}_H(\mathbf{H}_A)$ intensity values have earlier been defined in (20). The fused image is constructed by querying the predictive mean of the GP model whilst performing a normalisation as described below. Additionally, a constant mean value $\mu = 0.5$, is assumed over the image, this is justified because image pixels are observed to be continuous within the range of 0-1. Hence, the predictive mean in (16) becomes

$$\mathbf{A}_i = \mu + k(\mathbf{H}_A, \mathbf{L}_A) [k(\mathbf{L}_A, \mathbf{L}'_A) + \sigma_n^2 \mathbf{I}]^{-1} (\mathbf{I}_L(\mathbf{L}_i) - \mu) \quad (26)$$

where i represent the i th spectral band of the low-resolution image, and \mathbf{I} is an identity matrix equivalent to the number of pixels in the low-resolution image. This implies that we query the GP model by the number of spectral channels present in the low-resolution image. Finally, the new modality image \mathbf{A}_* is the concatenated sum of all \mathbf{A}_i 's.

C. Fusion Quality Metrics

It is important to evaluate the quality of the output fused image from the model using well-established image quality assessment measures. Methods such as Image Correlation Coefficient (CC) and Mean Squared Error (MSE) are widely used to determine the quality of images and measure the similarity between fused image and a reference image. However, image quality measures can be classified into three categories depending on the aim of the fusion. This can be to measure spatial, spectral or global quality of the image. In [32] for example, a Wald's Protocol is used to test the quality of pan sharpened images focusing on consistency and synthesis. In this paper, the following global quality performance measures are used to test the quality of the fused image

1. **Root Mean Squared Error (RMSE)**: RMSE evaluates the difference between the fusion model output and the reference image [33], providing a complete image quality pointer with results closer to zero indicating a high performance of the algorithm. RMSE is given as

$$RMSE(\tilde{\mathbf{R}}_*, \mathbf{A}_*) = \left(\frac{\sum_{i=1}^M \sum_{j=1}^N [\tilde{\mathbf{R}}_*(i, j) - \mathbf{A}_*(i, j)]^2}{M \times N} \right)^{1/2} \quad (27)$$

where $\mathbf{R}_*(i, j)$ and $\mathbf{A}_*(i, j)$ are the pixel values of the fused and reference image, respectively and $M \times N$ are the rows and columns of the image that define the image size.

2. **Image Correlation Coefficient (CC)**: CC is a fusion quality measure that characterises the geometric distortion between the reference image and the estimated (fused) image. The higher the correlation between the images the better the estimation of the spectral values. CC is defined as

$$CC(\mathbf{R}_*, \mathbf{A}_*) = \frac{\sum_{mn} (\mathbf{R}_* mn - \bar{\mathbf{R}}_*) (\mathbf{A}_* mn - \bar{\mathbf{A}}_*)}{(\sum_{mn} (\mathbf{R}_* - \bar{\mathbf{R}}_*)^2 (\sum_{mn} (\mathbf{A}_* - \bar{\mathbf{A}}_*)^2)^{1/2}} \quad (28)$$

where \mathbf{R}_* is a pixel of the reference image with size $(m \times n)$, $\bar{\mathbf{R}}_*$ is the mean of the reference image. Similarly, \mathbf{A}_* with size $(m \times n)$ represent a pixel of the fused image and $\bar{\mathbf{A}}_*$ is its mean.

4. **Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS)**: This measure offers a global indication of quality of the estimated image [2], based on normalised average error of each band of the image [34]. The ideal value of ERGAS is 0 [16]. Increase in the value of ERGAS could mean a distortion in the estimated image, on the hand increase in the value of ERGAS indicates that the estimated image $\mathbf{A}_*(x, y)$ is similar to the reference image $\tilde{\mathbf{R}}(x, y)$. ERGAS is defined as

$$ERGAS(\tilde{\mathbf{R}}, \mathbf{A}_*) = 100r \sqrt{\frac{1}{N} \sum_{j=1}^N \left(\frac{RMSE_j}{\mu_j} \right)^2} \quad (29)$$

where r is the ratio between the resolution of the images fused, μ_j is the sample mean of the i th band of \mathbf{A}_* .

III. SEGMENTATION

Segmentation is the subdivision of the image into separated regions [14], grouping similar pixels to homogeneous image segments so that increase in heterogeneity over the image is very much reduced allowing image pixels to be classified correctly in a decision oriented application. In oil spill detection, segmentation is a pre-requisite for classifying oil spill and look-alike. In this stage, dark areas that are assumed oil spill based on appearance are segmented out from the image and features are extracted that form the base for classification. In this paper, the segmentation phase is done using the K-means clustering algorithm.

A. K-means Segmentation

K-means clustering allows partitioning of data into a k number group of the data [35], classifying the given data (image) into k number of disjoint clusters. To achieve this, the algorithm is divided into two steps. In the first step, it calculates the k centroid in the image using Euclidean distance, and in the second step it groups each image pixel to a cluster nearest to a k_i centroid from the respective pixel. Summarily, K-means is an iterative method that minimizes all distances from each pixel to its cluster k_i centroid over all clusters k . Using the output image (\mathbf{A}_*) from (26), with size $m \times n \times \lambda$, where λ is the number of bands in \mathbf{A}_* . The aim is to segment the image into k number of clusters, let $a(m_i, n_j)$ be an input pixel of \mathbf{A}_* to be assigned to a cluster, and c_k be the centroid of the clusters; first the number of clusters k is initiated, secondly, for each pixel of the image, the Euclidean distance d is calculated, between the centre c_k of the centroid and the pixel using

$$d = \| a_*(m_i, n_j) - c_k \| \quad (30)$$

Next, all pixels are assigned to the nearest k_i using d . A new position of k_i is recalculated using

$$c_k = \frac{1}{k} \sum_{n \in c_k} \sum_{m \in c_k} a_*(m_i, n_j) \quad (31)$$

Finally, the cluster of pixels are reshaped into the segmented image.

IV. EXPERIMENTAL RESULTS

A. Dataset

The dataset used in this paper are as presented in Table 1. Firstly, multi-modal and multi temporal SAR images of the Gulf of Mexico oil spill as acquired by the Canadian RADARSAT-2 ScanSAR instrument are utilised. This instrument is fully polarimetric (HV,VV,VH) in wide beam mode with a nominal swath of 500km. Secondly, the European Space Agency (ESA)'s Envisat system with a single band (VV) polarisation also in ScanSAR and wide swath mode is also utilised.

TABLE I: Characteristics of the Dataset:

Satellite	Instrument	Resolution	Band	Dimension	Date Acquired
Radarsat-2	SAR	100m	C	865 × 905	29/04/10
Envisat	ASAR	150m	C	930 × 1271	26/04/10

B. Results and Discussion

The first task is to pre-process the SAR images to be fused, to reduce speckle noise and enhance the image using a Gaussian filter as discussed in Section II. The registration stage is next, following the steps described in Section II to align the images and to find correspondence between them. A mosaic of the two images is presented in [7] to show the progression of dark area (supposed spill location) over the multi-temporal period. An average fusion result is shown in Fig 4(a); a product of adding the two images together and taking the average. The GP fusion algorithm described in Section III is applied to fuse the multi resolution images. Subsequently, we compare the results of the proposed algorithm with [21] using global image quality measures described in Section III.D to test the quality of the output image and the performance of the proposed method. The results of validating the performance of the proposed algorithm is compared with other methods and presented in Table II. In Figs. 4b and 4c we present the results of the fusion process of the proposed method and the method of [21]. It is noticed that the proposed approach achieves compelling enhancement visually; this is attributed to the intrinsically sparse covariance function that provides a much smoother prediction of the function. In Fig. 5, the RMSE per image pixel of the output image \mathbf{A}_* and the output image from the method of [21] is also compared. Again, the proposed method achieves a better performance in this measure. The evaluation time to output the fused image from evaluating the covariance kernel function with optimised hyperparameter values is also compared. The proposed method achieved this in 0.2sec compared to [21] which took 0.63sec. Lastly, K-means technique described in Section III is applied to segment the dark formations (assumed oil spill) in \mathbf{A}_* by converting the image into *Lab* colour spaces which gives the initial value of K , as described in (30) and (31), respectively. The segmentation result is shown in Figs. 4d and 4e, objects in the image are clustered to different regions in Fig. 4f.

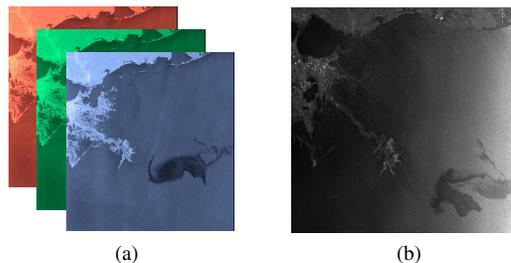


Fig. 3: RADARSAT-2 ScanSAR and Envisat ASAR images of Gulf of Mexico Oil Spill acquired 29/04/2010 and 26/04/2010, respectively.

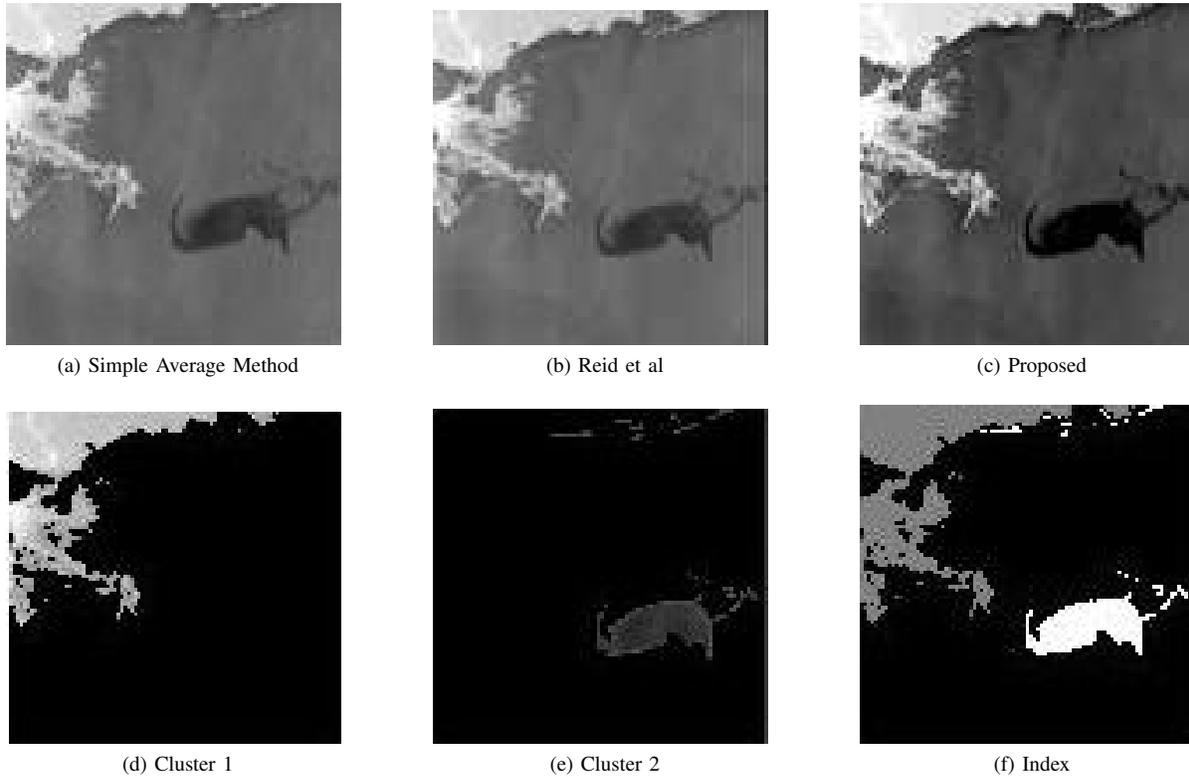


Fig. 4: (Top) Fusion Results: (a) Simple Average (b) Reid et.al(c) Proposed (bottom) Segmentation Result: (d) Cluster 1 (e) Cluster 2 (f) Index

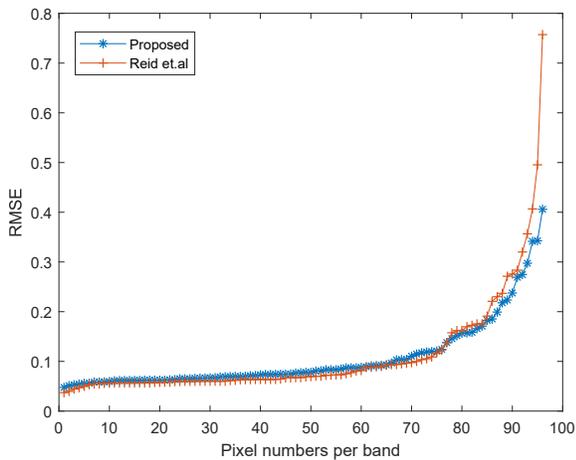


Fig. 5: Comparison of RMSE per pixel of Methods

TABLE II: Quality Measures of the Fusion Result

Fusion Method	CC	RMSE	ERGAS
DWT Fusion	0.6414	0.4937	23.3466
Reid et al	0.9997	0.7571	5.5657
Proposed	0.9978	0.4059	5.8099

V. CONCLUSIONS

This paper proposes a GP approach to fuse SAR images of different modalities, and to segment dark areas for oil spill detection. An automatic feature based image registration (SIFT) is utilised to find matching features and create correspondence between the multi-modal and multi-temporal images fused. The mosaic created using RANSAC algorithm has shown progression of the dark area signifying spread of oil spill when compared with the individual SAR images before fusion. This is useful in monitoring, and for further image based analysis, including for classification.

ACKNOWLEDGMENT

The authors would like to acknowledge Petroleum Trust Development Fund (PTDF) Nigeria, for funding this work.

REFERENCES

- [1] I. Leifer, W. J. Lehr, D. Simecek-Beatty, E. Bradley, R. Clark, P. Denison, Y. Hu, S. Matheson, C. E. Jones, B. Holt *et al.*, "State of the art satellite and airborne marine oil spill remote sensing: Application to the bp deepwater horizon oil spill," *Remote Sensing of Environment*, vol. 124, pp. 185–209, 2012.
- [2] L. Loncan, S. Fabre, L. B. Almeida, J. M. Bioucas-Dias, W. Liao, X. Briottet, G. A. Licciardi, J. Chanussot, M. Simoes, N. Dobigeon *et al.*, "Hyperspectral pansharpening: a review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 27–46, 2015.
- [3] J. W. Griggs, "Bp gulf of mexico oil spill," *Energy LJ*, vol. 32, p. 57, 2011.

- [4] M. Fingas and C. Brown, "Review of oil spill remote sensing," *Marine Pollution Bulletin*, vol. 83, no. 1, pp. 9–23, 2014.
- [5] C. Brekke and A. H. Solberg, "Oil spill detection by satellite remote sensing," *Remote sensing of environment*, vol. 95, no. 1, pp. 1–13, 2005.
- [6] D. H. S. Group, "Final report on the investigation of the Macondo well blowout," *Center for Catastrophic Risk Management, University of California at Berkeley*, 2011.
- [7] F. S. Longman, L. Mihaylova, and D. Coca, "Oil spill segmentation in fused synthetic aperture radar images," in *Proceedings of the 4th International Conference on Control Engineering & Information Technology (CEIT)*. IEEE, 2016, pp. 1–6.
- [8] M. Cococcioni, L. Corucci, and B. Lazzerini, "Issues and preliminary results in oil spill detection using optical remotely sensed images," in *Proceedings of OCEANS 2009-EUROPE*, 2009.
- [9] K. Topouzelis, V. Karathanassi, P. Pavlakis, and D. Rokos, "Detection and discrimination between oil spills and look-alike phenomena through neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 4, pp. 264–270, 2007.
- [10] K. N. Topouzelis, "Oil spill detection by sar images: dark formation detection, feature extraction and classification algorithms," *Sensors*, vol. 8, no. 10, pp. 6642–6659, 2008.
- [11] A. D. Lawal, G. Radice, M. Ceriotti, and A. U. Makarfi, "Investigating SAR algorithm for spaceborne interferometric oil spill detection," *International Journal of Engineering and Technical Research*, vol. 4, no. 3, pp. 123–127, 2016.
- [12] M. Marghany, "Automatic detection of oil spills in the Gulf of Mexico from RADARSAT-2 SAR satellite data," *Environmental Earth Sciences*, vol. 74, no. 7, pp. 5935–5947, 2015.
- [13] S. Singha, T. J. Bellerby, and O. Trieschmann, "Detection and classification of oil spill and look-alike spots from sar imagery using an artificial neural network," in *Proceedings of 2012 IEEE International conference on Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2012, pp. 5630–5633.
- [14] K. Topouzelis, V. Karathanassi, P. Pavlakis, and D. Rokos, "Oil spill detection: SAR multi-scale segmentation & object features evaluation," in *In Proceedings of the 9th International symposium on remote sensing (SPIE)*, 2002, pp. 77–87.
- [15] R. Chandrakanth, J. Saibaba, G. Varadan, P. A.-D. Raj *et al.*, "Fusion of high resolution satellite SAR and optical images," in *Proceedings of 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM)*. IEEE, 2011, pp. 1–6.
- [16] V. R. Pandit and R. Bhiwani, "Image fusion in remote sensing applications: A review," *International Journal of Computer Applications*, vol. 120, no. 10, 2015.
- [17] L. Dabbiro, S. Samiappan, R. A. A. Nobrega, J. A. Aanstoos, N. H. Younan, and R. J. Moorhead, "Fusion of synthetic aperture radar and hyperspectral imagery to detect impacts of oil spill in Gulf of Mexico," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 1901–1904.
- [18] R. Guida, S. W. Ng, and P. Iervolino, "S- and x-band sar data fusion," in *2015 IEEE 5th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, Sept 2015, pp. 578–581.
- [19] M. Berbar, S. Gaber, and N. Ismail, "Image fusion using multi-decomposition levels of discrete wavelet transform," in *Proceedings of International Conference on Visual Information Engineering, 2003. VIE 2003*. IET, 2003, pp. 294–297.
- [20] X. Yang, J.-H. Pei, and W. Yang, "Disadvantage of the methods based on wavelet transform in high-resolution and multispectral fusion image," *Journal of Infrared and Millimeter Waves*, vol. 21, no. 1, pp. 77–80, 2002.
- [21] A. Reid, F. Ramos, and S. Sukkarieh, "Bayesian fusion for multi-modal aerial images," in *Robotics: Science and Systems*, 2013.
- [22] C. M. Christoudias, R. Urtasun, M. Salzmann, and T. Darrell, "Learning to recognize objects from unseen modalities," in *European Conference on Computer Vision*. Springer, 2010, pp. 677–691.
- [23] M. Subramanyam and Mahesh, "Automatic feature based image registration using SIFT algorithm," in *Proceedings of 2012 Third International Conference on Computing Communication & Networking Technologies (ICCCNT)*. IEEE, 2012, pp. 1–5.
- [24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [25] M. Subramanyam and Mahesh, "Automatic feature based image registration using SIFT algorithm," in *Proceedings of 2012 Third International Conference on Computing Communication & Networking Technologies (ICCCNT)*. IEEE, 2012, pp. 1–5.
- [26] R. Chen, M. Hawes, L. Mihaylova, J. Xiao, and W. Liu, "Vehicle logo recognition by Spatial-SIFT combined with logistic regression," in *Proceedings of the International Conference on Information Fusion 2016*, 2016.
- [27] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. MIT press Cambridge, 2006, vol. 1.
- [28] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [29] H. He and W.-C. Siu, "Single image super-resolution using gaussian process regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 449–456.
- [30] A. Melkumyan and F. Ramos, "Multi-kernel gaussian processes," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, no. 1, 2011, p. 1408.
- [31] —, "A sparse covariance function for exact Gaussian process inference in large datasets," in *IJCAI*, vol. 9, 2009, pp. 1936–1942.
- [32] M. F. Yakhdani and A. Azizi, "Quality assessment of image fusion techniques for multisensor high resolution satellite images (case study: IRS-P5 and IRS-P6 satellite images)," in *Proceedings of the ISPRS TC VII Symposium-100 years ISPRS*, 2010.
- [33] V. R. Pandit and R. Bhiwani, "Image fusion in remote sensing applications: A review," *International Journal of Computer Applications*, vol. 120, no. 10, 2015.
- [34] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," *Aquatic Procedia*, vol. 4, pp. 133–142, 2015.
- [35] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm," *Procedia Computer Science*, vol. 54, pp. 764–771, 2015.