This is a repository copy of *Multi-band image fusion using Gaussian process regression with sparse rational quadratic kernel*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/146912/

Version: Accepted Version

Proceedings Paper:
Longman, F., Mihaylova, L. orcid.org/0000-0001-5856-2223, Yang, L. et al. (1 more author) (2020) Multi-band image fusion using Gaussian process regression with sparse rational quadratic kernel. In: 2019 22th International Conference on Information Fusion (FUSION). 22nd International Conference on Information Fusion, 02-05 Jul 2019, Ottawa, Canada. IEEE . ISBN 9781728118406

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works. Reproduced in accordance with the publisher's self-archiving policy.

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/

Multi-Band Image Fusion Using Gaussian Process Regression with Sparse Rational Quadratic Kernel

Fodio S. Longman, Lyudmila Mihaylova, Le Yang† and Konstantinos N. Topouzelis ‡
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@canterbury.ac.nz†, topouzelis@marine.aegean.gr ‡
Department of Electrical and Computer Engineering,
University of Canterbury, Chrischurch, New Zealand. †
Department of Marine Sciences,
University of the Aegean, University Hill, 81100, Mytilene, Greece. ‡

***Abstract*—This paper proposes an approach for multi-band image fusion using a multiple output variable Gaussian Process (GP) model. The considered model uses a new covariance function, which is a product of an intrinsically sparse kernel and a Rational Quadratic Kernel (RQK) to model the pixel coordinates and intensity of the high spatial resolution image. The new kernel serves as a stochastic prior for each band of the estimated image. The developed approach allows the exchange of information between the different modalities enabling local structure of the high spatial resolution image on which the model is trained. The accuracy performance and image quality assessment show that the proposed approach achieves compelling enhancement when compared with other fusion methods.**

Index Terms: Image Fusion, Remote Sensing, Gaussian Processes, Multi-output Variable Gaussian Processes

I. Introduction

Multi-sensor data fusion allows the integration of information originating from different sources with the aim of acquiring a unified information that benefits from complimentary information from different sensor data [1]. Image fusion is a subset of the diverse research area of multi sensor data fusion that provides the framework and tools that enable data originating from different imaging platforms to be aligned, producing an image of greater quality than individual ones. [2], [3].

Technological advancement have enabled diverse design of imaging sensors including hyperspectral, Synthetic Aperture Radar (SAR) and multi-spectral systems usually mounted on satellites, aircraft or Unmanned Aerial Vehicles (UAVs) to acquire multi modal image data, providing the scientific community with a wide range of data in different spatial and spectral resolutions. Efficient fusion methods are necessary to enable a comprehensive understanding of the scene of interest from the image and to improve the overall quality of images [4].

In remote sensing, a conventional image fusion approach is pansharpening that involves fusing a high spatial resolution panchromatic (PAN) image and a low spatial but high spectral resolution multispectral (MS) image to produce an image with high spatial and high spectral resolution [5]. However, algorithms developed for this purpose apply smoothness prior and rely on the sensitivity model of the sensor to obtain super resolution images while assuming common spectral channels [6]. The problem of fusing images in different modalities is an ill-posed one that needs regularization that can be achieved by fitting a suitable prior distribution on the image [7].

Gaussian Processes (GP) is a powerful data driven approach that led to efficient solutions of image processing problems including for modelling of low level image features, image de-noising and exploring the structural redundancy for super-resolution image reconstruction. GPs have been applied to high resolution object reconstruction for lower resolution images by considering local structures in natural images defined by their pixel neighbourhood [8]. Additionally, GP regression provides a flexible framework for fusing multiple datasets from heterogeneous sensors [9]. In [6] a multi-task GP approach is developed for multi-modal image fusion. Other image fusion approaches such as Wavelet Transform have been used in [10], [11], [12] to improve spatial resolution and maintain the spectral property of the images. This was achieved by converting the image from its spatial domain to the frequency domain, and decomposing the image into approximated and fine details without information loss. This fusion approach allows image properties to be inferred, thanks to the fusion rule including Haar, Debuchies and other wavelets [13].

This paper presents an approach that combines high resolution image data acquired by a UAV exploring marine environments with multispectral data acquired by the Sentinel-2 satellite for the same scene. This work extends the approach proposed in [13] for fusing images with multiple bands. Firstly, a multi-output regression framework is introduced by applying an independent GP [14] to each output band in the high resolution image. Secondly, common spectral channels in the visible range are selected from the multi-modal images (3-bands in each dataset). The covariance function of the model is the product of an intrinsically sparse kernel and a Rational Quadratic Kernel (RQK) utilised

to model the coordinates and intensities of pixels in the high resolution image which forms the spatial base of the estimated image to be recovered. This work is motivated by the ability of GP's to sufficiently model the relationship between different modalities with varied outputs, with an aim at extending its inherent point data operation over areas [6]. The contribution of this work is the adaptation of a multi-output variable GP to model the spatial information from selected bands of the high spatial resolution image and to use a mapping cross covariance to combine spectral information. The proposed kernel serve as a base stochastic prior for each band of the estimated image.

The rest of the paper is organised as follows. In section II, we make a quick review of GP and introduce multi-output variable GP. In section III, we describe in more detail the proposed approach including the prior covariance design and hyper-parameter estimation. In section IV, the dataset and experiment are described with simulated results and discussion. Section V, concludes the paper.

II. Gaussian Processes

A GP can be described by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ also known as the kernel. It is a stochastic process that defines a collection of random variables [15] . The covariance function is defined by some hyperparameters that characterise its behaviour while the mean function may conveniently assume a zero value since GP can be modified to model a non-zero mean [16]. More insight into different kernels and modelling the mean function can be found in [17]. Consider a function $f(\mathbf{x})$, whose mean and kernel are defined as

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

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

where $\mathbb{E}$ is the expectation operator and the GP is then described as a non-linear function expressed as:

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

The GP model can achieve the mapping of inputs $\mathbf{x}_i \in \mathbb{R}^{\mathcal{D}}$ to an output space $\mathbf{y}_i \in \mathbb{R}$ by imposing a Gaussian prior over the latent function where the output vector $\mathbf{y}$ is a noisy observation represented as

$$\mathbf{y} = f(\mathbf{x}) + \epsilon \tag{4}$$

and

$$\epsilon \sim \mathcal{N}(0, \sigma^2) \tag{5}$$

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

The parameters of its kernel are then learnt using N input-output pairs from a given training dataset $\mathcal{D}$, such that $\mathcal{D} = \{(\mathbf{x_1}, \mathbf{y_1}), \dots (\mathbf{x_N}, \mathbf{y_N})\}$. The function values are normally distributed with the mean $m(\mathbf{x})$ and the covariance $k(\mathbf{x}, \mathbf{x}')$ given by:

$$[f(\mathbf{x_1})^T f(\mathbf{x_2})^T \dots f(\mathbf{x_N})^T]^T = \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{6}$$

here, $m(\mathbf{x})$ and $k((\mathbf{x}, \mathbf{x}')$ are as defined in (1) and (2), respectively.

The entries of the covariance matrix $\mathbf{K}_{ij}$ are calculated by evaluating each element of $k(\mathbf{x}, \mathbf{x}')$ using the user defined kernel e.g the Squared Exponential (SE) with its hyperparameters [17]. This can be expressed as

$$k(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} k(\mathbf{x_1}, \mathbf{x_2}) & k(\mathbf{x_1}, \mathbf{x_2}) & \dots & k(\mathbf{x_1}, \mathbf{x_N}) \\ k(\mathbf{x_2}, \mathbf{x_1}) & k(\mathbf{x_2}, \mathbf{x_2}) & \dots & k(\mathbf{x_2}, \mathbf{x_N}) \\ \vdots & \vdots & \vdots & \vdots \\ k(\mathbf{x_N}, \mathbf{x_1}) & k(\mathbf{x_N}, \mathbf{x_2}) & \dots & k(\mathbf{x_N}, \mathbf{x_N}) \end{bmatrix} \tag{7}$$

As part of the GP training, the hyperparameters of the kernel are optimized. This is achieved by maximising over the hyperparameters and the marginal likelihood. This is illustrated in the next section. Consequently, the joint distribution of the training outputs $\mathbf{y}$ and the test outputs $\mathbf{y}^*$ with the mean function is defined by

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu} \begin{bmatrix} \mathbf{X} \\ \mathbf{X}^* \end{bmatrix}, \begin{bmatrix} K_{\mathbf{XX}} + \sigma_n^2 \mathbf{I} & K_{\mathbf{XX}^*} \\ K_{\mathbf{X}^*\mathbf{X}} & K_{\mathbf{X}^*\mathbf{X}^*} \end{bmatrix} \right) \tag{8}$$

In (8) above, $\mathbf{X}$ and $\mathbf{X}^*$ define the design matrices for the training and test data, respectively. When $\mathbf{y}^*$ is conditioned on the observations $\mathbf{y}$ the predictive distribution becomes

$$p(\mathbf{y}^*|\mathbf{X}, \mathbf{y}, \mathbf{X}^*) \sim \mathcal{N}(\boldsymbol{\mu_\star}, \boldsymbol{\Sigma_\star}) \tag{9}$$

where

$$\boldsymbol{\mu_\star} = k(\mathbf{X}^*, \mathbf{X}) K_X^{-1} \mathbf{y} \tag{10}$$

$$\boldsymbol{\Sigma_\star} = k(\mathbf{X}^*, \mathbf{X}^*) - k(\mathbf{X}^*, \mathbf{X}) K_X^{-1} k(\mathbf{X}, \mathbf{X}^*) \tag{11}$$

Here, $\boldsymbol{K}_X$ is defined as

$$\boldsymbol{K}_X = k(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 \mathbf{I} \tag{12}$$

and $\sigma_n^2 \mathbf{I}$ is the standard deviation of the measurement noise, $\mathbf{I}$ is an N-dimensional identity matrix. With the learnt hyperparameter values of the kernel, the GP can then predict the output $\mathbf{y}^*$ using the predictive distribution of input and outputs points.

A. Multi-Output Variable Gaussian Processes

Normally, a GP model assumes single output value. However, multiple output values are possible in practice. A common approach is to model each output value as an independent GP model [14]. Additionally, new kernels can be constructed over multi-dimensional inputs by adding or multiplying between kernels defined on each individual input. An additive function can simply be expressed as $f(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ and can easily be encoded into GP models. It allows a flexible way to model functions having more than one input. Consider the functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ drawn independently from a GP prior, for two different datasets $\mathcal{D}_1$ and $\mathcal{D}_2$,

$$f_1(\mathbf{x}) \sim \mathcal{GP}(0, k_1(\mathbf{x}, \mathbf{x}')) \tag{13}$$

$$f_2(\mathbf{x}) \sim \mathcal{GP}(0, k_2(\mathbf{x}, \mathbf{x}')) \tag{14}$$

where the input ouput data pairs is given by:

$$\mathcal{D}_1 = \{(\mathbf{x_1}, \mathbf{y_1}), \ldots (\mathbf{x_N}, \mathbf{y_N}\}) \tag{15}$$

and

$$\mathcal{D}_2 = \{(\mathbf{x_2}, \mathbf{y_2}), \ldots (\mathbf{x_N}, \mathbf{y_N}\}) \tag{16}$$

The distribution of the sum is given by

$$f_1(\mathbf{x}) + f_2(\mathbf{x}) \sim \mathcal{GP}(\mu_1 + \mu_2, \mathbf{K}_1 + \mathbf{K}_2) \tag{17}$$

where $\mathbf{K} = k(\mathbf{x}, \mathbf{x}')$. The above expression can be used to sum any number of components. Consequently, the modelling of functions of multiple dimensions will result to additive structure across the dimensions such that

$$f(\mathbf{x}, \mathbf{x}') \sim \mathcal{GP}(0, \mathbf{K}_1 \mathbf{K}_2) \tag{18}$$

where the individual covariances are defined by $\mathbf{K}_1$, $\mathbf{K}_2$ and assuming a 0 mean, respectively.

III. Proposed Approach

The approach in this paper is motivated by the works in [13], [6] and extends them to multi-band images from heterogeneous imaging sensors. The aim is to combine two image modalities with different spatial and spectral characteristics. The approach considers the change of support problem inherent in image fusion where pixels are of different resolution with their neighbours. The proposed approach is given in Fig 1.

Let $\mathcal{I}H_{m \times n \times \lambda}$ and $\mathcal{I}L_{m \times n \times \lambda}$ represent the high spatial and high spectral resolution images acquired by two heterogeneous sensors for the same scene $\mathcal{X}$. Here, $._{m \times n}$ defines the spatial and $._{\lambda}$ the spectral extent of the images. The total number of measurements ($\mathcal{T}_z$) in each image is $._{m \times n \times \lambda}$ dimensions for $\mathcal{I}H$ and $\mathcal{I}L$, respectively. The images to be fused represent debased versions of the image to be recovered. For simplicity, we select common spectral channels in the visible range composed of Red, Green and Blue (RGB) in the images to be fused which means $\lambda = 3$. The problems we aim to solve are 1) To reconstruct a high-spatial and high spectral image from the complimentary images $\mathcal{I}H$ and $\mathcal{I}L$; 2) To build a covariance kernel that can handle pixel inference where the change of support problem exist; 3) To calculate the cross covariance that models the multi-modal and multi-band images. For this, we introduce a multi-output Gaussian Processes to model each spatial band of the high spatial resolution image. This forms the spatial prior of the image to be recovered.

A. Covariance Function

The covariance function or kernel plays a fundamental role in a GP model. It has the ability to encode our assumption of the function we want to model, and defines the correlation between function values [8]. The kernel is characterized by its hyperparameters. A commonly used covariance function is the Squared Exponential (SE) with hyperparameters σ_f^2 and l. Conventionally, a covariance kernel is defined on $\mathbf{x} - \mathbf{x}'$ and explores the relationship between these points. However, when dealing with images the covariance function can be extended over areas by relating pixel observation to the function $\mathbf{f}(\mathbf{x})$, where a pixel observation is defined by $P_k(_{x_i, y_i, z})$, here $_{x_i, y_i}$ defines the geometric location of the ith pixel and $_z$ its intensity value at that location in the kth band of the multi-band image. A detailed derivation of defining covariance over areas of the image is given in [6].

$$\mathbf{P}(\mathbf{H_A}) = \frac{1}{|A_H|} \iint_{\mathbf{x} \in \mathbf{H_A}} f(x) dx \tag{19}$$

$\mathbf{H_A}$ is the geometry areas per band of the high spatial resolution pixels, and $\mathbf{H_L}$ is the low spatial and high spectral pixels, the aim is firstly, to design the prior that defines the structure of the high spatial and high spectral image to be recovered. For this, we model the high spatial resolution pixels of the high spatial image by defining areas, where an area consists of the intensity of the pixel and the geometry of it at that location. We assume a simple average relationship between pixels and the GP function. The covariance of two high spatial resolution pixels in the ith band is then defined as

$$k_B(\mathbf{H_A}, \mathbf{H'_A}) = \frac{1}{|\mathbf{H_A}||\mathbf{H'_A}|} \iint_{\mathbf{x} \in \mathbf{H_A}} \iint_{\mathbf{x}' \in \mathbf{H'_A}} k(\mathbf{x}, \mathbf{x}') dx dx' \tag{20}$$

where $k_B(\mathbf{H_A}, \mathbf{H'_A})$ defines the covariance between two high spatial resolution pixels in the ith band of the image and B ranges from 1-number of bands in the high spatial resolution image and $|\mathbf{H'_A}|$ is the surface area of $\mathbf{H_A}$. In designing the prior covariance function, the following factors are considered:

- Image data are normally non-smooth
- Image data exhibit discontinuity
- Spatial non-stationarity of images

The spatial information in $\mathcal{I}H$ is used as the input space of the prior covariance function added with the observed pixels to achieve contextual non-stationarity that addresses the discontinuity problem. Firstly, we explore an intrinsically sparse covariance introduced in [18] to enable sparsity and reduce computational complexity inherent in the full storage of the covariance matrix $\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}$. The proposed kernel is suitable for applications exhibiting discontinuities. It is smooth but not infinitely differentiable. Let $\mathbf{M_S}$ be the intrinsically sparse covariance defined by

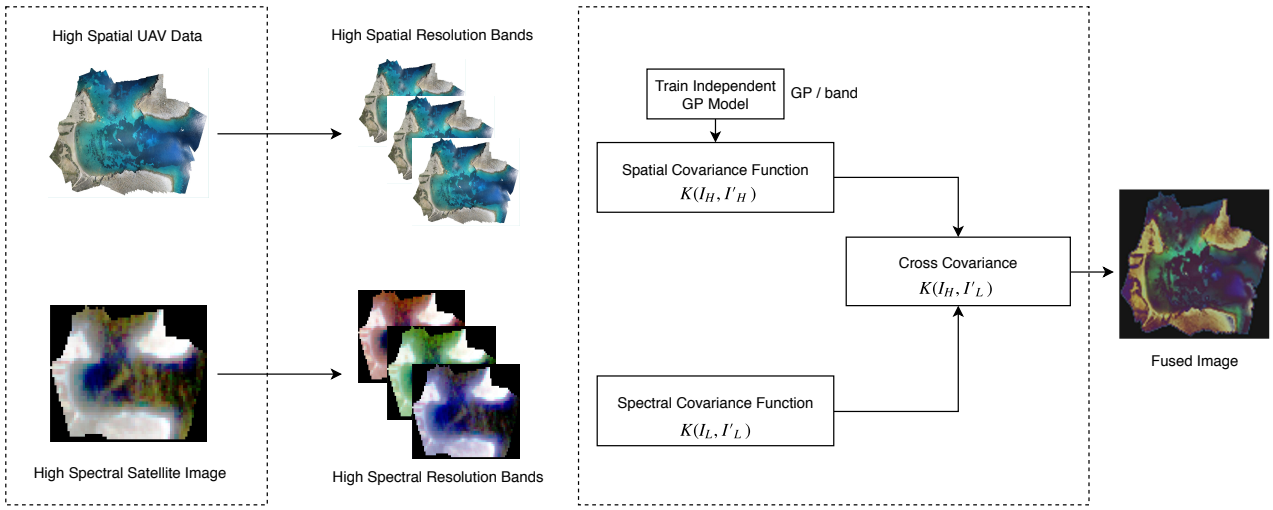


Fig. 1: Multi-Band Fusion Framework

$$\mathbf{MS}_B(\mathbf{H_A}, \mathbf{H'_A}; \sigma_0, l)_i =$$

$$\begin{cases} \sigma_0 \left[\frac{2+\cos(2\pi \frac{d}{l})}{3}(1-\frac{d}{l}) + \frac{1}{2\pi}\sin(2\pi \frac{d}{l}) \right] & \text{if } d < l \\ 0 & \text{if } d \geq l \end{cases} \tag{21}$$

The proposed kernel is a function of the midpoint coordinates of areas (pixel coordinates and the intensity value at that location) in the ith band and reduces to zero for displacements larger than d. The variables σ_0 and l are the hyperparameters of the kernel. Here $\sigma_0 > 0$ is a constant coefficient that determines the average distance of $\mathbf{f}(\mathbf{x})$ from μ_*, $l > 0$ is the characteristic length scale that determines the length of change of $\mathbf{f}(\mathbf{x})$, and d is defined by

$$\mathbf{d}_i = |\mathbf{H_A} - \mathbf{H'_A}| \tag{22}$$

The second step of the prior covariance design utilises the Rational Quadratic Kernel (RQK) [17] to model the pixel intensities of each band in the high spatial resolution image, allowing the pixels to be linked with the observed paired pixel geometries. The RQK is equal to putting together several SE kernels with different characteristic lengths that enables smoothness transfer. It improves efficiency in handling the change of support problem. Let $\mathbf{M_P}$ be the RQK kernel defined by

$$\begin{aligned} &\mathbf{MP}_B(\mathbf{I}(\mathbf{H_A}), \mathbf{I}(\mathbf{H'_A}); \alpha, \theta_p)_i \\ &= \sigma^2 \left(1 + \frac{(\mathbf{I}(\mathbf{H_A}) - \mathbf{I}(\mathbf{H'_A}))}{2\alpha\theta_P^2} \right)^{-\alpha} \end{aligned} \tag{23}$$

Here, $\mathbf{IH_A}, \mathbf{IH'_A}$ represent the pixel intensity values in the $\mathbf{H_A}$ area of the ith band of the high spatial resolution image. The role of $\mathbf{M_S}$ is to ensure sparsity that improves computation while $\mathbf{M_P}$ links pixels within the covariance using the multi-band spatial information from the high spatial resolution multi-band image. The hyperparameters of the RQK are θ_p and α, respectively. Lastly, the prior covariance can be defined as the product of two independent kernels that models the spatial and pixel intensity of the high spatial resolution image forming the spatial base of the estimated image. This is mathematically given by

$$k_B(\mathbf{H_A}, \mathbf{H'_A}) = \sigma^2 \mathbf{MS}((\mathbf{H_A}, \mathbf{H'_A}; \boldsymbol{\theta_1}) \mathbf{MP}(\mathbf{I}(\mathbf{H_A}), \mathbf{I}(\mathbf{H'_A}); \boldsymbol{\theta_2}) \tag{24}$$

The GP model is then trained using this kernel to learn the hyperparameters of the model here represented as θ_1 and θ_2 where k_B represents the number of spatial bands in the high spatial resolution image.This means an independent GP is trained for each spatial band of the high spatial image. The product prior function augments the geometry of pixels and the intensities. In the next section we describe the learning of the hyperparameters of the model.

B. Hyperparameter Optimization

The hyperparameters of the GP model refers to the parameters of the prior covariance function evaluated on each band of the multi- band high spatial resolution image. The parameters of the model include μ_*, $\sigma_n^2\mathbf{I}$. The hyperparameter vector associated with the covariance function is defined as $\theta = \{\sigma_0, \theta_p, \alpha, l\}$. In the proposed covariance function θ_p controls sensitivity and α determines the relative weighting for scale variations. When $\alpha \to \infty$, the behaviour of the RQK is similar to the SE kernel. It is important to initialise the parameters sensibly as they determine the quality of the estimated image to be recovered. A non-optimal solution is likely to produce fusion result that is blurry or even with high frequency artifacts [8].

To optimize θ, the marginal likelihood is maximised, which is given by:

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\mathbf{f}, \mathbf{X}) p(\mathbf{f}|\mathbf{X}) d\mathbf{f} \tag{25}$$

From (3) and (4), the likelihood $\mathbf{y}|\mathbf{f} \sim \mathcal{N}(\mathbf{f}, \sigma_n^2\mathbf{I})$ and the model prior over the latent function $\mathbf{f}$ gives the logarithm of the marginal likelihood

$$log(p(\mathbf{y}|\mathbf{X},\boldsymbol{\theta})) = -\frac{1}{2}\mathbf{y}^T K_X^{-1}\mathbf{y} - \frac{1}{2}log|K_X| - \frac{n}{2}log2\pi. \quad (26)$$

In (26), the first term finds data fit, the second term is the model complexity term and the third term is a constant that ensures the marginal likelihood is robust to over-fitting.

C. Cross Covariance

The optimized θ values from (24) are used to calculate the base covariance $k(\mathbf{H_A}, \mathbf{H'_A})$ that forms the spatial base of the estimated image to be recovered. While applying independent GP to multi-output regression problems is seen as a sub-optimal approach because the cross correlation is not put into consideration, the proposed approach solves this problem by mapping image areas where an area consists of both the pixel geometry and intensity values. $\mathbf{H_A}$ areas are discretely integrated by assuming they correspond to the $\mathbf{H_L}$ areas as shown in [6].

The optimized base covariance is then used to calculate the cross covariance $k(\mathbf{H_A}, \mathbf{H'_L})$ per band in $\mathbf{H_L}$. The role of this kernel is to couple per band the high spatial resolution pixels with the corresponding high spectral pixels. The cross covariance is given by

$$k_B(\mathbf{H_A}, \mathbf{L'_A}) = \frac{1}{\mathcal{T}_\mathcal{H}} \sum_{\mathbf{H'_A} \in \mathbf{L'_A}} k(\mathbf{H_A}, \mathbf{H'_A}) \quad (27)$$

where $\mathcal{T}_\mathcal{H}$ defines the total number of $\mathbf{H_A}$ areas correponding to $\mathbf{H_L}$ areas. The function is evaluated over the corresponding bands in $\mathbf{H_L}$. Similarly, the cross covariance between areas in $\mathbf{H_L}$ can be calculated using

$$k_B(\mathbf{L_A}, \mathbf{L'_A}) = \frac{1}{\mathcal{T}_\mathcal{H}\mathcal{T}'_\mathcal{H}} \sum_{\mathbf{H_A} \in \mathbf{L_A}} \sum_{\mathbf{H'_A} \in \mathbf{L'_A}} k(\mathbf{H_A}, \mathbf{H'_A}) \quad (28)$$

D. Image Fusion and Reconstruction

To fuse the image modalities and reconstruct the estimated image, the training data of the model is extracted from the high spatial resolution image $\mathbf{H_A}$ consisting of the spatial geometries of pixels and the augmented intensity values within the bands observed. The GP model is then queried over the high spectral image $\mathbf{H_L}$ pixel areas and their intensities. The reconstruction is done by querying the predictive mean of the model (see (9)-(11)), again we evaluate this function by querying corresponding bands in the modalities. A constant mean value of 0.5 is assumed because image values are continues within the range of 0-1. The predictive mean in (10) becomes

$$\mathcal{I}_B^* = \mu + k_B(\mathbf{H_A}, \mathbf{H_L}) \left[k_B(\mathbf{L_A}, \mathbf{L'_A}) + \sigma_n^2 \boldsymbol{I}\right]^{-1} (\mathbf{I_L}(\mathbf{L_i}) - \mu) \quad (29)$$

E. Fusion Performance Metric

It is important to evaluate the performance of the fusion model to validate the result. Conventionally, this is done by comparing the fusion results with a reference image using fusion performance metrics e.g. image Correlation Coefficient (CC) [11], Reconstruction Error (RE) or the Universal Image Quality Index (UIQI) [7]. In situations where a reference image is not available, non-reference image fusion metrics have been developed. We propose to use the Fast- Feature Mutual Information (Fast-FMI) introduced in [19] as a measure to validate the fusion model. Fast-FMI calculates the mutual information between corresponding regions in the fused and source images, respectively. Firstly, the mutual information is normalized using

$$\frac{\ell(\mathcal{I}_i) + \ell(\mathcal{I}^*)}{2} \quad (30)$$

Here, $\ell(\mathcal{I}_i)$ and $\ell(\mathcal{I}^*)$ defines the entropies of corresponding windows in the source images and the fused image, respectively. Secondly, the mutual information between the source images and the fused image is defined by

$$I(i;\mathcal{I}^*) = \frac{2}{n}\sum_{i=1}^{n} \frac{I_i(\mathcal{I}_i;\mathcal{I}^*)}{\ell(\mathcal{I}_i) + \ell(\mathcal{I}^*)} \quad (31)$$

Finally, the non-reference fusion metric is given by

$$FMI_{\mathcal{I}^*}^{\mathcal{I}_i} = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{I_i(\mathcal{I}_H;\mathcal{I}^*)}{\ell(\mathcal{I}_H) + \ell(\mathcal{I}^*)} + \frac{I_i(\mathcal{I}_L;\mathcal{I}^*)}{\ell(\mathcal{I}_L) + \ell(\mathcal{I}^*)}\right) \quad (32)$$

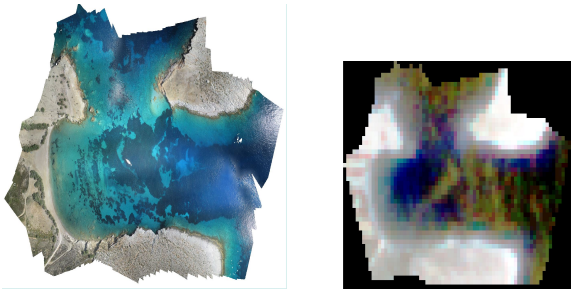
IV. Experiment

A. Dataset

In this section, we describe the dataset used for the experiment. We utilise multi-band data from a UAV and satellite data from the sentinel-2 satellite. While the UAV data is high in spatial resolution, it lacks spectral information. On the other hand, the satellite data has low spatial resolution and high spectral resolution. The aim of our model therefore, is to bring the complimentary information from the two heterogenous sensors to reconstruct a high spatial and high spectral image for the scene of interest. Sentinel-2 is a multispectral earth observation satellite and forms part of the European Union (EU) Copernicus Programme for environmental monitoring. It has 13 bands in the visible, near infrared and short wave infrared part of the spectrum, with 10m spatial resolution in the visible range comprising the Red, Green, Blue and Near Infrared bands.

B. Results and Discussion

This section evaluates the performance of the proposed Multi-Output variable GP model. The images are registered with pixel correspondence established between them. Image registration determines the geometric transformation that aligns one image to the other [20]. We propose to recover a high spatial and high spectral resolution image by fusing the two complimentary images. Firstly, the images are resized to 100×100 and normalized so that image pixels are in the



(a) (b)

Fig. 2: Quick Look of UAV and Sentinel-2 Multi Band Images

range of 0-1 and to reduce the computational complexity inherent in regression models. Secondly, common spectral bands in the visible range (Red, Green and Blue) are selected from the images. Pixel geometry per spatial band in $\mathcal{I}_{\mathcal{H}}$ and the corresponding intensity values have been extracted forming $\mathbf{H_A}$ and $\mathbf{I}(\mathbf{H_A})$, respectively. The variables extracted form the input of the prior function are shown in (19)-(24). The model is then trained using the proposed base kernel and the marginal likelihood is maximised to optimize the hyperparameters of the model. A cross covariance is then calculated using the optimized base kernel that couples the high spatial pixels and the corresponding high spectral bands per band using a mapping of 4:1 corresponding areas between $\mathbf{H_A}$ and $\mathbf{H_L}$, respectively. In Fig.3 we show the high spatial and high spectral resolution images with the fused image result using the considered model and their corresponding image histogram. Image histogram shows the graphical representation of pixel intensity on the x-axis and the corresponding number of pixels on the y-axis. The results shows that there are more pixels in the fused image when compared to the individual images, implying more information in the fused image.

In Fig.4 (top), we compare the proposed fusion approach with the simple averaging and inverse wavelet transform methods. Simple averaging, sums the source images and finds the average while the wavelet approach decomposes the images into approximations and fuses them using a rule. Both methods do not consider spatial location of pixels and their corresponding intensities while performing fusion. Our method on the other hand, uses both characteristics of the image as the input space to the model. This allows transferral of information between the image modalities that enables local structure of images to be learned. In the results, simple averaging shows ringing leading to artifacts and blurred image, while the inverse wavelet transform exhibits mosaic effects which could be due to small decomposition scale. The proposed method however, shows visually appealing image with improved edges that alleviates artifacts as can be seen on the top right corner of Fig.4.

Additionally, we test the performance of our method by segmenting the fused image using K-means clustering algorithm [13] and compare the result with the other methods. Again, our method shows compelling results, picking up more objects from the image than the other methods as shown in Fig. 4(bottom). We attribute this to the ability of the considered approach to sharpen the resolution of the spectral channels that supports pixel coordinates from the high spatial image improving the edges.

Finally, the Fast-FMI non-reference image fusion performance metric described in (30)-(32) is utilised to validate the performance of the proposed approach. The proposed method is compared with other established image fusion methods and subjected to the performance test. Visual comparisons in Fig.4 (top) shows the proposed method having visually appealing performance than the other methods. A window of 3×3 is used which corresponds to evaluating $(3 \times 3)^2$ regions and finding the MI between them. Fast-FMI performance metric reduces the computational complexity from $\mathcal{O}(n)^2$ to $\mathcal{O}(n)$ when compared to similar non-reference performance measures.

TABLE I: Fast-FMI Results

Method	Fast-FMI
Simple Averaging	0.4321
IWT	0.4532
Proposed	0.3954

V. Conclusion

In this paper, we propose the fusion of multi-band images acquired by heterogeneous imaging sensors using independent single-output Gaussian Processes to model spatial and spectral bands. A non-stationary product covariance function is utilised to model the geometry and intensities of pixels in the high spatial resolution image forming the spatial base of the fused image. A cross covariance is then calculated using the base kernel and a mapping corresponding areas between different modalities. The multi-output approach allows the exploitation of local image structure that improved sharp edges with little visual artifacts when compared with the other methods. The results of Fast-FMI shows the proposed approach performing better than the compared methods. We attribute this to the ability of the model to capture more information from the complimentary images by finding pixel correspondences and adapt to the change of resolution problem inherent in multi-band images.

Acknowledgment

The authors would like to acknowledge Petroleum Trust Development Fund (PTDF) Nigeria, for funding this work. We would also like to thank Professor Topouzelis Konstantinos of the University of the Aegean in Greece for providing the datasets.

References

[1] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[2] V. R. Pandit and R. Bhiwani, "Image fusion in remote sensing applications: A review," *International Journal of Computer Applications*, vol. 120, no. 10, 2015.

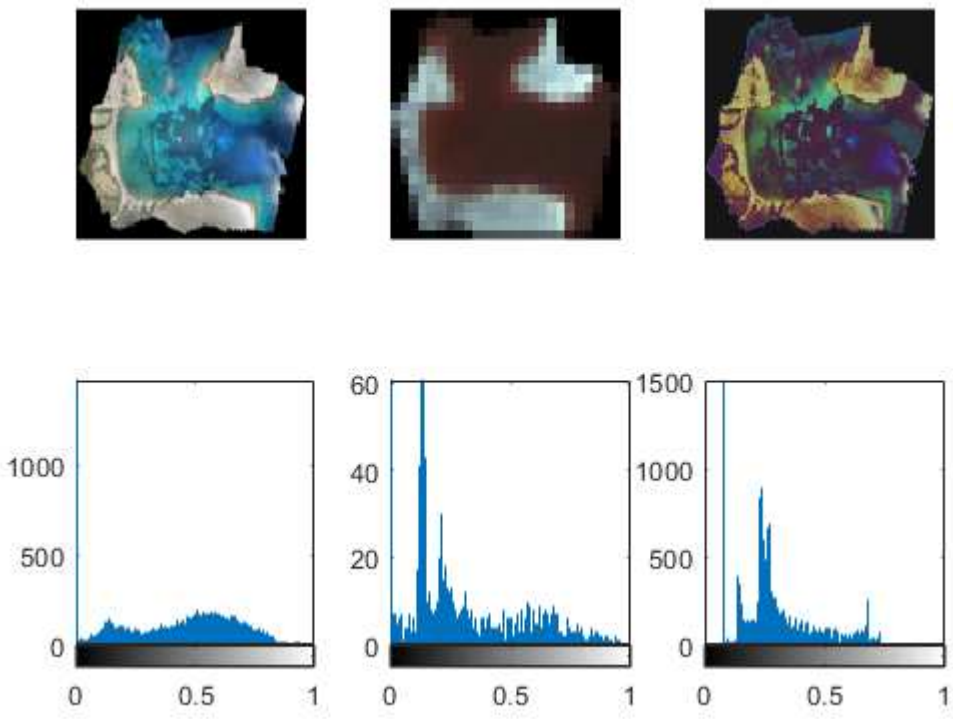
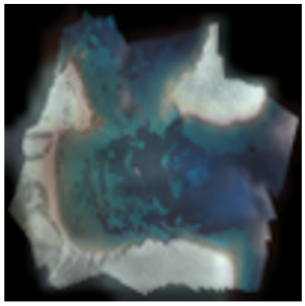
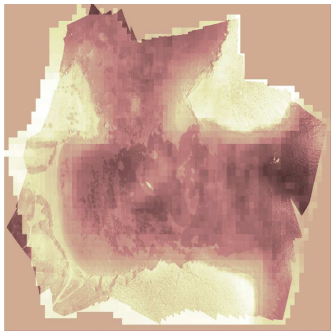


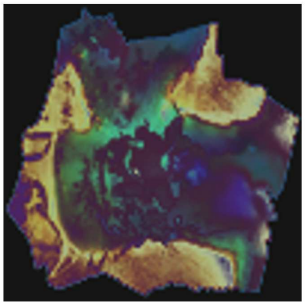
Fig. 3: (Top) left-right: UAV Image, Satellite Image, Proposed Result (below) Corresponding Histograms



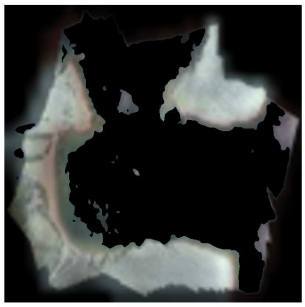
(a) Simple Average Method



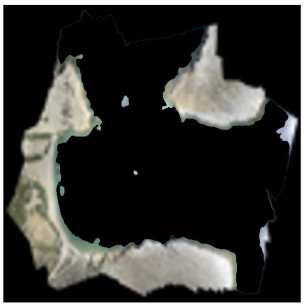
(b) IWT



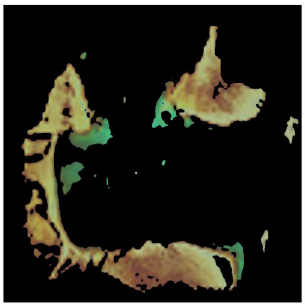
(c) Proposed



(d) Cluster 1



(e) Cluster 2



(f) Cluster 3

Fig. 4: (Top) Fusion Results: (a) Simple Average (b) Inverse Wavelet Transform (c) Proposed (bottom) Segmentation Result: (d) Cluster 1 (e) Cluster 2 (f) Cluster 3

[3] 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.

[4] V. PohlC, "Multisensorimagefusioninremotesensing: concepts, methodsandapplications," *InternationalJournalofRemoteSensing*, vol. 19, no. 5, p. 823r854, 1998.

[5] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3176–3180.

[6] A. Reid, F. Ramos, and S. Sukkarieh, "Bayesian fusion for multi-modal aerial images." in *Robotics: Science and Systems*, 2013.

[7] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multiband images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1117–1127, 2015.

[8] 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.

[9] S. Vasudevan, "Data fusion with Gaussian processes," *Robotics and Autonomous Systems*, vol. 60, no. 12, pp. 1528–1544, 2012.

[10] 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.

[11] 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.

[12] 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.

[13] F. S. Longman, L. Mihaylova, and L. Yang, "A Gaussian process regresion approach for fusion of remote sensing images for oil spill segmentation," in *proceedings of 2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 62–69.

[14] L. Yang, K. Wang, and L. S. Mihaylova, "Online sparse multi-output gaussian process regression and learning," *IEEE Transactions on Signal and Information Processing over Networks*, 2018.

[15] W. Aftab, A. De Freitas, M. Arvaneh, and L. Mihaylova, "A Gaussian process approach for extended object tracking with random shapes and for dealing with intractable likelihoods," in *Digital Signal Processing (DSP), 2017 22nd International Conference on*. IEEE, 2017, pp. 1–5.

[16] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[17] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. MIT press Cambridge, 2006, vol. 1.

[18] A. Melkumyan and F. Ramos, "A sparse covariance function for exact Gaussian process inference in large datasets." in *IJCAI*, vol. 9, 2009, pp. 1936–1942.

[19] M. Haghighat and M. A. Razian, "Fast-FMI: non-reference image fusion metric," in *Application of Information and Communication Technologies (AICT), 2014 IEEE 8th International Conference on*. IEEE, 2014, pp. 1–3.

[20] 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.