JOINT RANGE ESTIMATION AND SPECTRAL CLASSIFICATION FOR 3D SCENE RECONSTRUCTION USING MULTISPECTRAL LIDAR WAVEFORMS

Yoann Altmann, Aurora Maccarone, Aongus McCarthy, Gerald Buller* and Steve McLaughlin†

School of Engineering, Heriot-Watt University, Edinburgh, U.K.
Email: {Y.Altmann, am827, A.McCarthy, G.S.Buller, S.McLaughlin}@hw.ac.uk

ABSTRACT
This paper presents a new Bayesian classification method to analyse remote scenes sensed via multispectral Lidar measurements. To a first approximation, each Lidar waveform mainly consists of the temporal signature of the observed target, which depends on the wavelength of the laser source considered and which is corrupted by Poisson noise. By sensing the scene at several wavelengths, we expect a more accurate target range estimation and a more efficient spectral analysis of the scene. Thanks to its spectral classification capability, the proposed hierarchical Bayesian model, coupled with an efficient Markov chain Monte Carlo algorithm, allows the estimation of depth images together with reflectivity-based scene segmentation images. The proposed methodology is illustrated via experiments conducted with real multispectral Lidar data.

Index Terms— Multispectral Lidar, Depth imaging, Bayesian estimation, Markov Chain Monte Carlo, Classification.

1. INTRODUCTION
Laser altimetry (or Lidar) is an acknowledged tool for extracting spatial structures from three-dimensional (3D) scenes. Using time-of-flight to create a distance profile, signal analysis can recover, for instance, tree and canopy heights, leaf area indices and ground slope by analyzing the reflected photons from a target. Conversely, passive multispectral (MSI) (dozen of wavelengths) and hyperspectral images (HSI) (hundreds of wavelengths) are widely used to extract spectral information about the scene which, for forest monitoring, can also provide useful parameters about the canopy composition and health. The most natural evolution to extract spatial and spectral information from sensed scenes is to couple Lidar data and multi/hyperspectral images [1, 2]. Multispectral Lidar (MSL) has recently received attention from the remote sensing community for its ability to extract both structural and spectral information from 3D scenes [3–5]. The key advantage of MSL is the ability to provide information on the full 3D distribution of materials, especially for scenes including distributed objects (e.g., vegetation). Another motivation for MSL (instead of Lidar/HSI fusion) is that HSI, even when fully synchronized, can only integrate the spectral response along the path of each optical ray, not measure the spectral response as a function of distance, e.g. depth into a forest canopy. In [5–7], spectral unmixing techniques have been developed to analyze 3D scenes composed of multi-layered objects, assuming that the spectral signatures of the materials composing the scenes were known and assuming linear mixing processes. In this paper we assume that for each pixel, the photons emitted by the pulsed laser sources at different wavelength are reflected onto a single surface. This is typically the case for short to mid-range (up to dozens of meters) depth imaging where the divergence of the laser source(s) can be neglected. Moreover, for such applications, the size of the laser beam projected onto a surface of the scene is generally relatively small yielding a high probability for this surface to be composed of a single material. In the absence or lack of spectral mixtures, it seems more sensible to use classification methods to analyse MSL data. This is precisely the aim of this paper which studies, for the first time, a spectral classification algorithm for the analysis of MSL data.

In contrast with the classical additive Gaussian noise assumption used for passive hyperspectral images, Poisson noise models are more appropriate for MSL signals. Indeed, Lidar and thus MSL systems usually record histograms of time delays between emitted laser pulses and the detected photons arrivals. Within each histogram bin, the number of detected photons follows a discrete distribution which can be approximated by a Poisson distribution due to the particle nature of light. Due to the design of the proposed experiments (performed indoor) and to simplify the estimation problem, we assume that the ambient light (e.g., solar illumination) and dark counts can be neglected. In this paper, we demonstrate the possibility of efficient 3D scene analysis by exploiting geometric and spectral information contained in MSL data, under favorable observation conditions. However, the proposed method can be easily extended to more difficult observation conditions, as discussed in the conclusions of the paper.

Adopting a Bayesian approach, appropriate prior distributions are chosen for the unknown parameters of the model considered here, i.e., the surface positions, the classification labels and each class parameters. The joint posterior distribution of these parameters is then derived and a Markov chain Monte Carlo (MCMC) method is used to generate samples according to the posterior of interest. This fully Bayesian approach allows a careful study of the estimation performance (through the derivation of measures of uncertainty). Although very interesting, algorithmic improvement in terms of computational complexity (e.g., using optimization methods) is out of scope of this paper and should deserve more efforts in future work.

The remainder of the paper is organized as follows. Section 2 introduces the observation model associated with MSL returns for a single-layered object to be analyzed. Section 3 presents the hierarchical Bayesian model associated with the classification problem considered and the associated posterior distribution. Section 4 describes the MCMC method used to sample from the posterior of interest and subsequently approximate appropriate Bayesian estimators. Results of experiments conducted on real MSL data are shown and discussed in Section 5 and conclusions are finally reported in Section 6.

*Part of this work was supported by the EPSRC via grant EP/N003446/1, EP/K01538/1, EP/M01326X/1
†Part of this work was supported by the EPSRC via grant EP/J015180/1.
2. PROBLEM FORMULATION

This section introduces the observation statistical model associated with MSL returns for a single-layered object which will be used in Section 3 for spectral classification of MSL data. Precisely, we consider a 4-D arrays $Y$ of Lidar waveforms and of dimension $N_{\text{row}} \times N_{\text{col}} \times L \times T$, where $N_{\text{row}}$ and $N_{\text{col}}$ stands for the number of rows and columns of the regular spatial sampling grid (in the transverse plane), $L$ is the number of spectral bands or wavelengths used to reconstruct the scene and $T$ is the number of temporal (corresponding to range) bins. Let $y_{i,j,t} = [Y]^T_{i,j,t} = [y_{i,j,t,1}, \ldots, y_{i,j,t,T}]^T$ be the Lidar waveform obtained in the pixel $(i,j)$ using the $t$th wavelength. The element $y_{i,j,t} = \{0,1\}$ is the photon count within the $t$th bin of the $t$th spatial band considered. Let $d_{i,j}$ be the position of an object surface at a given range from the sensor, whose spectral signature (observed at $L$ wavelengths) is denoted as $\lambda_{i,j} = [\lambda_{i,j,1}, \ldots, \lambda_{i,j,L}]^T$. According to [8, 9] and assuming that the ambient illumination and dark photon counts can be neglected, each photon count $y_{i,j,t}$ is assumed to be drawn from the following Poisson distribution

$$y_{i,j,t}|\lambda_{i,j}, t_{i,j} \sim \mathcal{P}(\lambda_{i,j}g_{0,t_0}(t - t_{i,j})) \quad (1)$$

where $g_{0,t_0}()$ is the photon impulse response whose shape can differ between wavelength channels and $t_{i,j}$ is the characteristic time-of-flight of photons emitted by a pulsed laser source and reaching the detector after being reflected onto a target at range $d_{i,j}$ ($d_{i,j}$ and $t_{i,j}$ are linearly related in free-space propagation). Moreover, the impulse responses $\{g_{0,t_0}()\}$ are assumed to be known and can usually be estimated during the imaging system calibration.

In this work, we further assume that the spectral signatures of the scene surfaces belong to an unknown set of at most $K$ distinct surfaces $\{\mu_k\}_{k=1}^K$, i.e., $\lambda_{i,j} \in \{\mu_k\}_{k=1}^K$ where $K$ is a user-defined parameter. Note that to physical considerations the unknown spectral signatures $\{\mu_k\}_{k=1}^K$ are assumed to be positive.

The problem addressed in this paper consists of jointly estimating the range of the targets (for all the image pixels), of identifying the scene surfaces belong to an unknown set of at most $K$ distinct surfaces $\{\mu_k\}_{k=1}^K$, i.e., $\lambda_{i,j} \in \{\mu_k\}_{k=1}^K$. According to [8, 9] and assuming that the ambient illumination and dark photon counts can be neglected, each photon count $y_{i,j,t}$ is assumed to be drawn from the following Poisson distribution

$$y_{i,j,t}|\lambda_{i,j}, t_{i,j} \sim \mathcal{P}(\lambda_{i,j}g_{0,t_0}(t - t_{i,j})) \quad (1)$$

where $g_{0,t_0}()$ is the photon impulse response whose shape can differ between wavelength channels and $t_{i,j}$ is the characteristic time-of-flight of photons emitted by a pulsed laser source and reaching the detector after being reflected onto a target at range $d_{i,j}$ ($d_{i,j}$ and $t_{i,j}$ are linearly related in free-space propagation). Moreover, the impulse responses $\{g_{0,t_0}()\}$ are assumed to be known and can usually be estimated during the imaging system calibration.

In this work, we further assume that the spectral signatures of the scene surfaces belong to an unknown set of at most $K$ distinct surfaces $\{\mu_k\}_{k=1}^K$, i.e., $\lambda_{i,j} \in \{\mu_k\}_{k=1}^K$ where $K$ is a user-defined parameter. Note that to physical considerations the unknown spectral signatures $\{\mu_k\}_{k=1}^K$ are assumed to be positive.

The problem addressed in this paper consists of jointly estimating the range of the targets (for all the image pixels), of identifying the scene surfaces belong to an unknown set of at most $K$ distinct surfaces $\{\mu_k\}_{k=1}^K$, i.e., $\lambda_{i,j} \in \{\mu_k\}_{k=1}^K$ where $K$ is a user-defined parameter. Note that to physical considerations the unknown spectral signatures $\{\mu_k\}_{k=1}^K$ are assumed to be positive.

3. BAYESIAN MODEL

3.1. Likelihood

Assuming that the MSL waveforms $y_{i,j} = \{y_{i,j,t}\}_{t=1}^T$ of a given pixel $(i,j)$ result from the photons reflection onto a surface associated with the spectrum $\mu_k = [\mu_k, \ldots, \mu_k]_L^T$ (i.e., the pixel $(i,j)$ is associated with the $k$th class), the likelihood associated with the pixel $(i,j)$ can be expressed as

$$f(y_{i,j}|z_{i,j} = k, \mu_k, t_{i,j}) = \prod_{t} f_P(y_{i,j,t}; \mu_k, g_{0,t_0}(t - t_{i,j})) \quad (2)$$

where $z_{i,j} \in \{1, \ldots, K\}$ is a discrete classification label and $f_P(\cdot; \mu)$ denotes the probability mass function of the Poisson distribution whose mean is $\mu$. Assuming that the detected photon counts/noise realizations, conditioned on their mean in all pixels are independent, the joint likelihood can be expressed as

$$f(Y|Z, M, T) = \prod_{i,j} f(y_{i,j}|z_{i,j} = k, \mu_k, t_{i,j}), \quad (3)$$

where $Z$ and $T$ are $N_{\text{row}} \times N_{\text{col}}$ matrices gathering the pixel labels and target ranges respectively and $M = \{\mu_1, \ldots, \mu_K\}$ gathers the $K$ unknown spectra.

3.2. Prior distributions

In this work, we do not account for the potential spatial correlations between the target distances in neighbouring pixels of the scene. Each target position is considered as a discrete variable defined on $T = \{t_{\min}, \ldots, t_{\max}\}$, such that $1 \leq t_{\min} \leq t_{\max} \leq T$ (in this paper we set $(t_{\min}, t_{\max}) = (1, T)$) and assign the target ranges the following independent prior

$$p(t_{i,j} = t) = \frac{1}{T}, \quad t \in T \quad (4)$$

where $T' = \text{card}(T)$. Note that more informative priors could be used instead of (4) to capture potential spatial correlations affecting the range profiles, as in [9] where a total-variation (TV) regularisation [10, 11] promoting piecewise constant depth image was used. However, and as will be illustrated in Section 5, when the number of spectral bands $L$ considered and the number of detected photon are significant, the depth estimation does not require informative regularization (as the $L$ bands are used to estimate $t_{i,j}$). For this reason and for paper length constraints, we simply consider the uniform priors (4).

When prior information about the $K$ spectral signatures in $M$ is available, it can be introduced through an appropriate prior model. Here, we assume that limited knowledge is available (fixed number of classes $K$) and we assign the matrix $M$ the following prior model

$$\mu_k | r, \eta \sim \mathcal{G}(\mu_k, r; \eta), \quad \forall k, r \quad (5a)$$

$$r \sim \mathcal{G}(r; r_0, c_0), \quad (5b)$$

$$\eta \sim IG(\eta; \epsilon, \epsilon), \quad (5c)$$

where $\mathcal{G}$ and $IG$ stand for gamma and inverse-gamma distributions and where $(r_0, c_0)$ and $\epsilon$ are fixed hyperparameters which specify the shape of the prior $f(r)$ and $f(\eta)$, respectively. Here we choose $(r_0, c_0) = (1, 10)$ and $\epsilon = 10^{-2}$, yielding weakly informative priors. The hierarchical model (5) assumes that all reflectivity parameters share a priori the same statistical properties (through $r$ and $\eta$) while being flexible enough to allow to wide range of positive spectral signatures $M$. Here you propose to estimate $(r, \eta)$ assuming that we do not have informative prior knowledge about these parameters. Of course, Eqs. (5b) and (5b) can be adapted if additional information is available. Note that it is also possible to arbitrarily fix $(r, \eta)$. However, this can be challenging in practice as these parameters controls the prior mean and variance (usually unknown) of the different classes.

Due to the natural geometric structures of natural scenes, we can often expect neighbouring surfaces of objects to represent the same spectral signatures. To model this prior belief, we consider a Markov random field as a prior distribution for each label $z_{i,j}$ given its neighbors $Z_{\mathcal{V}(i,j)}$, i.e., $f(z_{i,j}|Z_{\mathcal{V}(i,j)}) = f(z_{i,j}|Z_{\mathcal{N}(i,j)})$, where $\mathcal{V}(i,j)$ is the neighborhood of the pixel $(i,j)$ and $Z_{\mathcal{N}(i,j)} = \{z_{i',j'}\}_{(i',j') \neq (i,j)}$. More precisely, we consider a Potts-Markov...
model which been wide used for multidimensional image segmentation (see [12, 13] among others). Given a discrete random field $Z$ attached to an image with $N$ pixels, the Hammersley-Clifford theorem yields $f(Z) = \frac{1}{G(\beta)} \exp \left[ \beta \sum_{i,j} \sum_{i',j' \in V(i,j)} \delta(z_{i,j} - z_{i',j'}) \right]$, where $\beta > 0$ is the granularity coefficient, $G(\beta)$ is a normalizing (or partition) constant and $\delta(\cdot)$ is the Dirac delta function. While several neighborhood structures can be employed to define $V(i,j)$, the eight pixels structure (or 2-order neighbourhood) will be considered in the rest of the paper.

3.3. Joint Posterior distribution

From the joint likelihood and prior model specified in Sections 3.1 and 3.2, we can now derive the joint posterior distribution for $T$, $M$, $Z$, $r$ and $\eta$ given the observed waveforms $Y$ and the value of the fixed hyperparameters $\Phi = (r_0, c_0, \eta)$. Using Bayes’ theorem, and assuming prior independence between $T$, $M$ and $Z$, the joint posterior distribution associated with the proposed Bayesian model is given by

$$f(T, M, Z, r, \eta | Y, \Phi) \propto f(Y | T, M, Z) f(T) f(M | \Phi) f(Z | r, \eta | \Phi)$$

(6)

4. ESTIMATION STRATEGY

The posterior distribution (6) models our complete knowledge about the unknowns given the observed data and the prior information available. To perform joint depth estimation and spectral classification of the MSL data, we use the following three Bayesian estimators: 1) the minimum mean square error estimator (MMSE) of the reflectivity matrix $\hat{M}_{\text{MMSE}} = E[M|Y, \Phi]$, the maximum a posteriori (MAP) estimator of target ranges $\hat{t}_{i,j,\text{MAP}} = \text{argmax}_{t_{i,j}} f(t_{i,j} | Y, \Phi)$ and 3) the MAP estimator of the classification labels $\hat{z}_{i,j,\text{MAP}} = \text{argmax}_{z_{i,j}} f(z_{i,j} | Y, \Phi)$. Note that we use the classical MAP estimators for the target ranges and pixel labels, as this estimator is particularly adapted to estimate discrete parameters. In order to approximate these estimators of interest, we adopt a fully Bayesian approach and consider a Markov chain Monte Carlo method to generate samples according to the joint posterior (6). More precisely, we use a Metropolis-within-Gibbs sampler to generate sequentially the unknown parameters from their conditional distributions and the samples are then used to approximate the Bayesian estimators of interest (after having discarded the first samples associated with the burn-in period of the sampler). The remainder of this section details the main steps of the proposed sampling strategy.

4.1. Sampling the target ranges

Using $f(T) = \prod_{i,j} p(t_{i,j})$, it can be seen from (3-4) and (6) that

$$f(T|M, Z, Y) \propto \prod_{i,j} f(t_{i,j} | M, Z, Y), \quad \forall t_{i,j} \in T. \tag{7}$$

Consequently, sampling the target ranges can be achieved by sampling independently each depth parameter from its conditional distribution, i.e., by drawing randomly from $\{t_{\min}, \ldots, t_{\max}\}$ with known probabilities.

4.2. Sampling the classification labels

In a similar fashion to the target ranges, sampling each label $z_{i,j}$ from its conditional distribution can be achieved by drawing in $\{1, K\}$ with known probabilities. In our experiments we used a Gibbs sampler implemented using a colouring scheme such that many labels can be updated in parallel (9 sequential steps required when considering a 2-order neighbourhood structure as each pixel has 8 direct neighbours).

4.3. Sampling the spectral signatures $M$

By exploiting the conjugacy of (3) and (5a) and

$$f(M | T, Z, Y, r, \eta) \propto \prod_{t, k} f(\mu_{t,k} | T, Z, Y, r, \eta), \tag{8}$$

sampling the $K$ independent gamma distributions with known parameters.

4.4. Sampling the hyperparameters $\tau$ and $\eta$

Using the conjugacy of (5a) and (5c), the conditional distribution of $\eta$ is a standard inverse-gamma distribution. Conversely, the conditional distribution of $\tau$ is a non-standard distribution and a Metropolis-Hastings step with Gaussian random walk proposal is used here to update $\tau$, in a similar fashion to [14]. Note that $f(\tau | Y, \eta, \Phi)$ is log-concave with $c_0 r_0 > 1$.

There are several Bayesian strategies for selecting the value of the regularisation parameter $\beta$ in a fully automatic manner (see [15] for a recent detailed survey on this topic). In this paper we use the empirical Bayes technique recently proposed in [16], where the value of $\beta$ is calculated by maximum marginal likelihood estimation. For brevity however, we assume $\beta$ is fixed in the remainder of this paper.

5. RESULTS

We propose assessing the performance of the proposed method to analyze the depth and spectral profiles of an approximately $5 \times 5$cm scene (see Fig. 1 (a)) composed of different objects made of Fimo clay and mounted on a painted backboard at a distance of 1.8m from a time-of-flight scanning sensor, based on TCSPC. The transceiver system and data acquisition hardware used for this work is broadly similar to that described in [9, 17–21], which was previously developed at Heriot-Watt University. The measurements have been performed indoor, in the dark to limit the influence of ambient illumination. The scene has been scanned using a regular spatial grid of $180 \times 180$ pixels and $L = 33$ regularly spaced wavelengths ranging from 500nm to 820nm. The histograms consist of $T = 3000$ bins of 2ps, which represents a depth resolution of $300\mu$m per bin. The power of the supercontinuum laser source has been adjusted from preliminary runs and the per-pixel acquisition time is 10ms for each wavelength. The instrumental impulse responses $g_{\alpha,\beta}(\cdot)$ were estimated from preliminary experiments by analyzing the distribution of photons reflected onto a Spectralon panel (a commercially available Lambertian scatterer). The proposed algorithm has been applied with $N_{\text{MC}} = 3000$ sampler iterations (including $N_{\text{b}} = 300$ burn-in iterations). Although spectral classification using state-of-the-art methods is possible after having estimated the target ranges, such approaches are by construction sub-optimal and are not discussed here due to space constraints. Fig. 1 (b) and (c) depicts examples of classification and range images, estimated with $K = 30$ classes.
The results show that in addition to the objects range estimation, it is possible to discriminate spectrally different surfaces (such as the different shades of green) and identify the regions of the scene where a given material is present (e.g., the red and blue clays). It is interesting to mention that the classification method is also able to identify spectral changes due to the orientation of the surfaces, as can be seen for the spherical yellow, blue and beige clays and around the corners of the other objects. Note that when increasing $K$, more spectral classes are associated with the backboard whose colour variation is much more significant than that of the other objects.

Finally, Fig. 3 compares the estimated spectra of the most visually similar objects of the scene, namely, the different green clays and the green backboard. This figure clearly illustrate the benefits of using MSL data for 3D scene analysis. First, by using a wide range of wavelengths ($L = 33$ here), it becomes possible to discriminate materials that would look similar using standard Lidar or a reduced number of wavelengths. For instance the green clays #2 and #3 present similar reflectivities between 600 and 650nm but more significant changes below 600nm. Second, using several wavelengths also helps estimating the depth profile. For instance, the green clay #5 and the green paint present low reflectivity coefficients between 600 and 700nm. Consequently, the number of detected photons at these wavelengths and associated with these objects are much lower than at longer wavelengths (for a given acquisition time and laser power), which can jeopardize the depth estimation. By splitting the energy of the laser source(s) across several wavelengths, the imaging system becomes more robust to objects that are hardly visible (low reflectivity) at particular wavelengths, which can be useful to reduce acquisition times while preserving imaging performance.

For completeness, Fig. 2 shows a joint range/spectral classification representation of the 3D-scene including only the 18 main classes of interests.

We proposed a new Bayesian model and a joint depth estimation and spectral classification algorithm for 3D scene analysis from MSL data. The scene surfaces visible by the imaging system were clustered into groups sharing the same spectral signatures. Adopting a Bayesian approach, prior distributions were assigned to the unknown model parameters; in particular, a Potts model was used to model the spatial organization of surfaces in natural scenes. Including ambient illumination and dark count levels in the observation model (as in [9, 21–23]) is the obvious next step from a more general application (especially for long-range imaging applications) of the proposed method. This paper shows that surface orientation clearly impacts the estimated object spectral response. In future work, it would be interesting to account for such spectral variations, as well as spectral variability of natural materials and the presence of distributed (multi-layered) targets. Although range-based scene segmentation is out of scope of this paper, coupling spectral and geometric information from the scene (e.g., for multi-layer classification) is a clearly interesting problem.

### 6. CONCLUSION

We proposed a new Bayesian model and a joint depth estimation and spectral classification algorithm for 3D scene analysis from MSL data. The scene surfaces visible by the imaging system were clustered into groups sharing the same spectral signatures. Adopting a Bayesian approach, prior distributions were assigned to the unknown model parameters; in particular, a Potts model was used to model the spatial organization of surfaces in natural scenes. Including ambient illumination and dark count levels in the observation model (as in [9, 21–23]) is the obvious next step from a more general application (especially for long-range imaging applications) of the proposed method. This paper shows that surface orientation clearly impacts the estimated object spectral response. In future work, it would be interesting to account for such spectral variations, as well as spectral variability of natural materials and the presence of distributed (multi-layered) targets. Although range-based scene segmentation is out of scope of this paper, coupling spectral and geometric information from the scene (e.g., for multi-layer classification) is a clearly interesting problem.
7. REFERENCES


