Spectral classification of sparse photon depth images

Altmann, Yoann; Maccarone, Aurora; McCarthy, Aongus; McLaughlin, Stephen; Buller, Gerald Stuart

Published in:
Optics Express

DOI:
10.1364/OE.26.005514

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Heriot-Watt University Research Portal

Citation for published version (APA):
Spectral classification of sparse photon depth images

YOANN ALTmann,† AURORA MACCARONE, AONGUS MCCARTHY, STEPHEN MCLAUGHLIN, AND GERALD S. BULLER

1Institute of Sensors, Signals and Systems, School of Engineering & Physical Sciences, Heriot-Watt University, EH14 4AS Edinburgh, UK
2Institute of Photonics and Quantum Sciences, School of Engineering & Physical Sciences, Heriot-Watt University, EH14 4AS Edinburgh, UK
†Y.Altmann@hw.ac.uk

Abstract: By illumination of target scenes using a set of different wavelengths, we demonstrate color classification of scenes, as well as depth estimation, in photon-starved images. The spectral signatures are classified with a new advanced statistical image processing method from measurements of the same scene, in this case using combinations of 33, 16, 8 or 4 different wavelengths in the range 500 - 820 nm. This approach makes it possible to perform color classification and depth estimation on images containing as few as one photon per pixel, on average. Compared to single wavelength imaging, this approach improves target discrimination by extracting more spectral information, which, in turn, improves the depth estimation since this approach is robust to changes in target reflectivity. We demonstrate color classification and depth profiling of complex targets at average signal levels as low as 1.0 photons per pixel from as few as 4 different wavelength measurements.

© 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

OCIS codes: (100.3010) Image reconstruction techniques; (030.5260) Photon counting; (330.1690) Color; (010.3640) Lidar.

References and links
1. Introduction

Color classification of multi-spectral images has significance in a number of applications, particularly in terms of remote mapping of terrain [1], underwater imaging [2], time-resolved fluorescence spectroscopy [3, 4] and target identification [5, 6], among others. This paper identifies new methods of color classification based on using a minimum of different spectral measurements and in terms of very low return signals, in the region of less than one photon per pixel, on average. The extraction of such detailed target information from sparse photon measurements is very challenging, but illustrates the use of photonics in cases of rapid information acquisition or extremely weak optical target return. This is of particular importance where low signal and rapid measurement are unavoidable, such as measurements from mobile platforms, e.g. remote sensing from moving airborne platforms [7], or undersea remotely operated and autonomous vehicles [8]. As well as maximizing the use of available sparse photon information, such approaches will result in considerably lower requirements for on-board data storage and data transmission bandwidth to base stations. A rapidly expanding area of interest is in using picosecond time-correlated single-photon detection for time-of-flight depth imaging. This approach has been used for a number of applications, for example kilometer-range depth imaging at centimeter resolution for target identification [6, 9], airborne mapping [7], and underwater imaging [10]. The advantages of shot-noise limited sensitivity and excellent timing resolution can provide significant potential for increased target range and/or reduced optical power requirements, as well as improved surface-to-surface resolution compared with more conventional time-of-flight lidar approaches. In recent years, there has been considerable effort in applying advanced image processing approaches to single-photon time-of-flight imaging. Recent studies, such as the first photon approach [11] and further adaptations [12], or variations of Bayesian methods [13, 14] have clearly demonstrated image depth and reflectivity reconstruction from sparse photon counts using a single wavelength.
Meanwhile, analysis of multi-wavelength single-photon data has also been investigated to enhance 3D scene characterization. In [15] and [16], we investigated a spectral clustering and a spectral classification method for multispectral Lidar data, respectively. While the clustering technique in [15] gathers within a same class those pixels exhibiting similar spectral signatures (and thus allow for intra-class spectral variability), the classification in [16] associates pixels that share the same spectral response. Spectral unmixing techniques, allowing the quantification of materials in 3D scenes have also been investigated [17,18]. Whilst these methods assume that all the wavelengths of interest are acquired at each pixel location, it is also possible to recover depth and color profiles from subsampled multispectral Lidar data by accounting for spatial dependencies between neighboring pixels [19]. This paper examines, for the first time, color classification from sparse multi-wavelength single-photon data and shows that such classification can be achieved, even with detected signal levels in the very low photon count regime. In a similar fashion to [16], we assume that pixels gathered in a given class share the same spectral response. Although accounting for spectral variability is interesting, it makes the segmentation problem particularly ill-posed in the sparse photon count regime. Indeed, subtle spectral variations become very difficult to detect and quantify and this effect is neglected in this work. To handle low photon count levels, we extend the model considered in [16] to further regularize the classification problem by promoting local smoothness for the depth profile.

For these initial studies we considered a depth imaging system which provides range and reflectivity information at different wavelengths (33, 16, 8 or 4) in the range 500 - 820 nm. The recorded data are then combined to classify the spectral responses of the scene and recover the corresponding range profile. Compared to single wavelength imaging, this approach improves target discrimination by extracting more spectral information, which, in turn, improves the depth estimation since this approach is robust to changes in target reflectivity. We proposed a computational method for joint depth profiling and spectral classification. Precisely, we adopted a Bayesian approach which allows the combination of observed data and additional a priori information in a formalized framework. Unfortunately, the resulting so-called posterior distribution associated with this challenging problem is highly multimodal and thus particularly difficult to exploit using standard optimization techniques. To tackle this problem, we used an efficient Markov chain Monte Carlo (MCMC) method in order to estimate the parameters of interest (depth profile and color classification map) and derive associated posterior measures of uncertainty (confidence intervals). By decoupling spectral classification and ranging, which is performed subsequently, the original problem can be decomposed into two successive simpler problems, which allows the use of a fast algorithm adapted to process large images and a large number of timing histogram bins.

2. Experimental setup

The depth imaging system was based on the time of flight approach and the time-correlated single-photon counting (TCSPC) technique. This technique measures the difference in time between a synchronization signal and a detected event, and stores this information in a timing histogram, as the measurement is repeated over many pulses. The experimental setup used to perform the depth profiles reported in this paper is shown schematically in Fig. 1. The system comprised a custom-built monostatic transceiver scanning unit, fiber-coupled to a supercontinuum laser source (SuperK EXTREME - EXW12, NKT Photonics, Denmark) used in conjunction with an acousto-optic tunable filter to select a single operational wavelength from the supercontinuum spectrum. The light from the laser system was delivered to the transmit channel of the transceiver unit via an armored photonic crystal fiber with 5 µm diameter core. The optical configuration of the transceiver unit was the same as described in [10], with coaxial transmit and receive channels, and a pair of computer controlled galvameter mirrors to raster scan the target. The optical components of the transceiver unit were chosen in order to optimize the system operation in
the wavelength range 500 - 820 nm. The transmitted light was focused on the target with a 100 mm focal length Canon EF series camera lens, with the aperture set to f/8. The light scattered by the target was collected by the objective lens and coupled back to the receive channel. The return signal was coupled into a 50 µm diameter core optical fiber connected to a Micro Photon Devices (MPD) PDM series silicon-based single-photon avalanche diode (Si-SPAD) detector. The detector had an active area of 50 µm diameter and a timing jitter of 50 ps full width at half maximum (FWHM). However, the overall timing jitter of the system was observed to vary between 50 ps and 100 ps FWHM, depending on the operational wavelength, as a result of spectral-dependent variations in laser pulse duration. The increased jitter contribution from the laser source at longer wavelengths was confirmed by independent streak camera measurements by the manufacturer. The use of a monostatic transceiver meant that significant back-reflections from the common transmit/receive channel could be detected by the SPAD. In order to avoid these back-reflections, we used an electrical gating approach to de-activate the detector at their expected arrival time. The supercontinuum laser had a repetition rate of 19.5 MHz and provided the synchronization signal to trigger the electrical gating and the TCSPC module (HydraHarp 400, PicoQuant GmbH, Germany). The TCSPC module time-stamped the laser synchronisation signal and the detected photon events. This information was transferred from the TCSPC module to the desktop computer via a USB 2.0 connection. Timing histograms, with 2 ps wide timing bins, were constructed for each pixel.

Several measurements were performed using the targets shown in Fig. 2, placed at a stand-off distance of approximately 1.85 meters from the transceiver unit. The target in Fig. 2(a) was made of several green clay objects and two leaves fixed to a dark grey backboard, and the target
Fig. 2. Photographs of targets: (a) shows Target 1 which contains green clay objects and natural materials, fixed to a backplane; and (b) shows Target 2 which was composed of 14 clay materials with different colours. The recorded images consist of $200 \times 200$ pixels (the scanned target areas were approximately $50 \times 50$ mm) and the targets were placed at 1.85 meters from the system.

in Fig. 2(b) was composed of 14 coloured clay objects fixed to a similar grey backboard. In both cases, the operational wavelength was varied with a step size of 10 nm in the wavelength range 500 - 820 nm, resulting in a total of 33 scans for each target. All the measurements were performed in a dark laboratory using an average optical power of approximately 480 nW. An area of approximately $50 \times 50$ mm at the target was scanned with $200 \times 200$ pixels (0.25 mm between adjacent pixels, vertically and horizontally), using an acquisition time of 10 ms per pixel. The performance of the proposed method has been assessed using subsets of the recorded data so that the average photon count for each band is 1 or 10. Precisely, the acquisition time has been artificially reduced, differently for each band, so that the photon counts averaged over all the image pixels are 1 or 10 for each band. For instance Fig. 3 shows examples of distributions of photon counts for an average of 1 photon per pixel and per band (Target 2). Note that in this case, about 45% of the pixels do not contain any detected photons. The diameter of the illuminating spot on the target was about 0.3 mm. For each wavelength, pixel-wise instrumental responses were measured by acquiring data with the same per-pixel acquisition time (10 ms). The target used to record the instrumental responses was a reference scatterer, a Spectralon panel (SRT-10-020 Spectralon Diffuse Reflectance Target, Labsphere) placed at the same standoff distance as the target, i.e., 1.85 meters from the system. However, due to the limited size of the reference scatterer (slightly smaller than $50 \times 50$ mm), we only analysed the central region of the scene, composed of $190 \times 190$ pixels.

3. Proposed computational method

As mentioned above, recovering spectral signatures from extremely photon-limited data is a challenging ill-posed problem which needs to be regularized to reduce estimation uncertainty and improve estimation/detection performance. Here we incorporate additional constraints/prior information via two discrete Markov random fields (MRFs) which capture the intrinsic spatial correlation of natural scenes. More precisely, for the range profile we used a Total-Variation (TV) based MRF [13], which promotes piecewise constant profiles. In a similar fashion to [16], a Potts model was adopted to promote homogeneous regions sharing similar spectral responses. If information about the expected spectral signatures of the scene is available (e.g. the materials present), it can also be included within the classification process. Here, to reflect the lack of prior knowledge about the objects of the scene, a classical weakly informative prior model is considered for the spectral signatures of the K spectral classes. Selecting the optimal number of
spectral classes $K$ in order to obtain a satisfactory trade-off between classification sensitivity and robustness (regarding spectral variability of natural objects) is a challenging application and scene-dependent problem. In this case, the number of classes is fixed a-priori. Solving the spectral classification problem, together with estimating the range profile is particularly difficult due to the multimodal nature of the likelihood associated with the classical single-photon Lidar observation model relying on Poisson noise assumption [13]. To alleviate convergence issues, we use an MCMC algorithm to generate random variables according to the posterior distribution obtained by coupling the likelihood with the proposed prior model. The generated samples are then used to estimate the range profile and the classification map. Assuming that the background levels (ambient illumination and detector dark counts) can be neglected, this allows the spectral classification problem to be solved first (as if a non-ranging single-photon imaging system was used), the range profile being estimated subsequently, using the spectral responses estimated. An important property of the proposed Bayesian approach is that the regularisation parameters of the MRFs are adjusted automatically, thus relieving the practitioner to tune them, e.g. via cross-validation.

3.1. Observation model

This section introduces the observation statistical model associated with multispectral Lidar (MSL) returns for a single-layered object which will be used for spectral classification of sparse single-photon Lidar data. Precisely, we consider a set of Lidar waveforms $Y$ of dimension $N \times L \times T$, where $N$ stands for the number of pixels associated with a 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 $y_{n,\ell,t} = [y_{n,\ell,t}]_{n=1}^{N} = [y_{n,\ell,t}]_{t=1}^{T}$ be the Lidar waveform obtained in the $n$th pixel using the $\ell$th wavelength. The element $y_{n,\ell,t}$ is the photon count within the $r$th bin of the $\ell$th spectral band considered. Let $d_n$ be the position of an object surface at a given range from the sensor, whose
spectral signature (observed at \( L \) wavelengths) is denoted as \( \lambda_n = [\lambda_{n,1}, \ldots, \lambda_{n,L}]^T \). Assuming that the ambient illumination and dark photon counts can be neglected, each photon count \( y_{n,t} \) is assumed to be drawn from the following Poisson distribution

\[
y_{n,t} | \lambda_{n,t}, t_n \sim \mathcal{P}(\lambda_{n,t} g_{n,t}(t - t_n))
\]  

(1)

where \( g_{n,t}(\cdot) \) is the photon impulse response whose shape can differ between pixels and wavelength channels and \( t_n \) is the characteristic time-of-flight of photons emitted by a pulsed laser source and reaching the detector after being reflected from a target at range \( d_n \) (\( d_n \) and \( t_n \) are linearly related in free-space propagation). Note that for most long-range applications using scanning systems, it is widely acknowledged that the impulse responses can be considered as spatially constant, i.e., \( g_{n,t}(\cdot) = g_t(\cdot), \forall n \). However, the observation model in (1), is able to account for inhomogeneous sensor sensitivity/efficiency (potentially observed in detector arrays) and potential limitations of the optical system, for example a non-uniform illumination or detection profile from the target. The impulse responses \( \{g_{n,t}(\cdot)\} \) are assumed to be known as they can be estimated with arbitrary precision during the imaging system calibration. We further assume that the spectral signatures of the scene surfaces belong to an unknown set of \( K \) distinct positive signatures \( \{\mu_k\}_{k=1,\ldots,K} \), i.e., \( \lambda_n \in \{\mu_k\}_{k=1,\ldots,K} \) where \( K \) is a user-defined parameter. In other words, we can introduce discrete classification labels \( z_n \in \{1, \ldots, K\} \) such that \( z_n = k \) if and only if the \( k \)th spectral signature is observed in the \( n \)th pixel (and \( \lambda_n = \mu_k \)). The problem of spectral classification considered in this work consists of estimating the classification labels \( \{z_n\} \) as well as the \( K \) spectral signatures of the \( K \) classes and the unknown target ranges. Note however that the proposed method can also be applied if the target ranges and/or the signatures of the spectral classes are known. To address this unsupervised classification problem, we adopt a Bayesian approach similar to that proposed in [16]. As discussed in the remainder of this section, the Bayesian framework allows us to incorporate prior information in order to regularize the problem and reduce estimation uncertainty. A Monte Carlo algorithm is then used to efficiently exploit the Bayesian model of interest.

3.2. Bayesian model

Assuming that the detected photon counts/noise realizations, conditioned on their mean in all pixels, spectral bands and time bins are independent, the joint likelihood function of the observed data can be expressed as

\[
f(Y|z, M, t) = \prod_n f(y_n|z_n = k, \mu_k, t_n),
\]  

(2)

where \( M = [\mu_1, \ldots, \mu_K] \) gathers the \( K \) unknown spectra and \( z \) and \( t \) are vectors of length \( N \) which gather the pixel labels and target ranges respectively.

In contrast to [16], we propose to account for the potential spatial correlations between the target distances in neighboring pixels of the scene. As will be shown in Section 4, this enables a more robust range estimation, in particular when the number of detected photons in each pixel is extremely low. Each target position is considered as a discrete variable and we define a prior model \( f(t|c) \) for the range profile which promotes spatial correlation between target ranges in neighbouring pixels. As in [13] we consider a total-variation (TV) based prior model [20,21] which promotes piece-wise constant range profiles. The amount of correlation between adjacent pixels is controlled by the regularization parameter \( c \) which is automatically tuned during the classification process.

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 \( M \) a weakly informative gamma/inverse-gamma hierarchical prior model. This model is denoted as \( f(M, \theta) \), where \( \theta \) gathers auxiliary parameters.
estimated during the estimation process. Moreover, this model is flexible enough to allow to wide range of positive spectral signatures \( \mathbf{M} \). Finally, a classical Potts-Markov model is used to define a prior model for the classification labels \( \mathbf{z} \). The spatial correlation induced by the resulting model \( f(\mathbf{z}|\beta) \) is controlled by the parameter \( \beta \) which is automatically adjusted (in a similar fashion to the regularization parameter \( c \)).

From the joint likelihood (2) and prior models \( f(\mathbf{t}|c) \), \( f(\mathbf{M}, \theta) \) and \( f(\mathbf{z}|\beta) \) we can derive the joint posterior distribution for \( \mathbf{t}, \mathbf{M}, \mathbf{z} \) and \( \theta \) given the observed waveforms \( \mathbf{Y} \) and the value of the regularization parameters \( \Phi = (c, \beta) \). Using Bayes’ theorem, and assuming prior independence between \( \mathbf{t}, \mathbf{M} \) and \((\mathbf{z}, \theta)\), the joint posterior distribution associated with the proposed Bayesian model is given by

\[
  f(\mathbf{t}, \mathbf{M}, \mathbf{z}, \theta|\mathbf{Y}, \Phi) \propto f(\mathbf{Y}|\mathbf{t}, \mathbf{M}, \mathbf{z}) f(\mathbf{t}|c) f(\mathbf{M}|\theta) f(\mathbf{z}|\beta).
\]

(3)

### 3.3. Estimation strategy

The posterior distribution (3) 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 \( \hat{\mathbf{M}}_{\text{MMSE}} = \mathbb{E}[\mathbf{M}|\mathbf{Y}, \Phi] \), 2) the marginal maximum a posteriori (MMAP) estimator of the classification labels \( \hat{\mathbf{z}}_{\text{MMAP}} = \arg\max f(\mathbf{z}_n|\mathbf{Y}, \Phi) \) and 3) the conditional maximum a posteriori (CMAP) estimator of target ranges \( \hat{t}_n|\text{CMAP} = \arg\max f(t_n|\mathbf{Y}, \hat{\mathbf{z}}_{\text{MMAP}}, \hat{\mathbf{M}}_{\text{MMSE}}, \Phi) \).

Note that we use the MMAP and CMAP estimators for the target ranges and pixel labels, as these estimators are particularly adapted to estimate discrete parameters. In order to approximate these estimators of interest, we adopt a Monte Carlo simulation approach and generate samples according to the joint posterior (3). 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). By restricting the admissible target ranges so that the target returns/peaks are not truncated, it is possible to decouple the ranging and spectral analysis problems. Precisely, the classification problem can be solved first by integrating the observations over their temporal dimension, which also reduces the computational complexity of the algorithm (in particular when \( T \) is large), then the range estimation is subsequently performed. One of the main advantages of Monte Carlo methods is that they often allow estimating the appropriate amount of regularisation from data. Indeed, there are several Bayesian strategies for selecting the value of the regularisation parameters \((c, \beta)\) in a fully automatic manner (see [22] for a recent detailed survey on this topic). Here we use the empirical Bayes technique recently proposed in [23], where the value of \((c, \beta)\) is calculated by maximum marginal likelihood estimation.

Thanks to its ability to estimate the regularisation parameters, the proposed method only relies on a single main user-defined parameter, namely the number spectral classes \( K \). Although the classification results depend on the values of \( K \) and on the spectral diversity/richness of the observed scene, slightly overestimating the value of \( K \) generally provides satisfactorily results. Indeed, the Potts-Markov model tends to identify homogeneous regions in the scene and tends to provide empty or scarcely represented classes when \( K \) is overestimated. Such classes can then be discarded during a post-processing step. Here we removed classes containing less than 1\% of the pixels. The corresponding pixels are then grouped as unclassified pixels. Note that it is also possible to group classes based on their spectral similarity following the proposed estimation process.
4. Results

This section presents spectral classification and depth estimation results obtained using the two targets whose images, including referenced objects, are presented in Fig. 2.

4.1. Spectral classification

Fig. 4. Analysis of Target 1 shown in Fig. 2 (a). The first column depicts the estimated depth profile (in mm), the reference range being arbitrarily set to the backplane range and the scale being the set to match those used in Fig. 11. The second depicts color classification maps. The third column depicts the spectral signatures of the most prominent classes, projected onto the first and second axes obtained using PCA. Each of these subplots illustrates the similarity between the estimated spectral signatures. Rows a) to d) depict results obtained with 33, 16, 8 and 4 wavelengths, regularly spaced within the 500-820 nm range (with an average of 10 detected photons per pixel, for each spectral band).

Figure 4 presents depth estimation and spectral classification results obtained associated with Target 1, composed of several clay objects and tree leaves and for different number of spectral bands. The right-hand side column is obtained by projecting the estimated signatures onto the two most significant components obtained by principal component analysis (PCA) [24]. This linear dimensionality reduction technique projects sequentially the original spectral signatures onto their principal subspace, spanned by orthonormal vectors (denoted by $b_1, b_2, \ldots \in \mathbb{R}^L$), which are ordered so that the first components $\tilde{\mu}_{k,i} = \mu_k^T b_i$ retain most of the variation present in all the original signatures. More precisely, in the right-hand side columns of Fig. 4,
we depict the unitless coefficients associated with the first (\{\tilde{\mu}_{k,1}\}) and second (\{\tilde{\mu}_{k,2}\}) principal components. The closer the coloured dots in Fig. 4, the more similar the spectral responses. As the number of spectral bands decreases, it becomes more difficult to detect subtle spectral variations (e.g. the fine leaf veins). However, it is still possible to discriminate the different materials composing Target 1 from measurements of only 4 wavelengths (Fig. 4(d)). Although the range estimation performance decreases with the number of wavelengths considered (as the overall number of detected photons decreases), it is possible to discriminate groups of materials presenting significant spectral difference, e.g. the leaves from the man-made objects (see right column of Fig. 4). In the left-hand side column of Fig. 4, it can be seen that pixels located at objects boundaries are considered as unclassified pixels (class colored in crystal blue in all the classification maps presented in this paper). Significant changes of the incident angle occur in these regions, which attenuates the amplitude of the measured reflectivity coefficients. Since such changes occur only in few pixels, the associated classes contain very few pixels and these classes are merged into a single unclassified class.

In addition to the overall number of spectral bands considered, the discrimination performance also depends on the number of detected photons in each pixel (or the per-pixel acquisition time) and the sparsity level of the spatial sampling, i.e. the total number of observed pixels. Figure 5
illustrates the colour classification of a $48 \times 48$ array from the same measurement, where only every fourth linear pixel is sampled, and all others are discarded. Figure 5 shows that reducing the pixel acquisition time may degrade the ability to discriminate spectrally similar materials such as the leaf structures. Similarly, reducing the spatial sampling resolution reduces the correlation between adjacent pixels and limits the detection of fine spatial structures. This figure shows that with 8 spectral bands, it is possible to discriminate the 6 types of man-made objects (i.e., materials #3 to #7 and material #9 in Fig. 2(a)). However, when considering only 4 bands and when the data quality reduces, the algorithm is no longer able to discriminate the materials #4 and #9.

![Image](image_url)

Fig. 6. Examples of estimated spectral signatures for the classes associated with the materials #4 (red) and #9 (blue) of the Target 1, with 33 wavelengths. The vertical dashed lines represent the selected wavelengths when considering only 4 spectral bands. The solid lines represent the posterior mean while the coloured regions represent confidence intervals of $\pm 3$ standard deviations around the posterior mean.

Figure 6 illustrates how the uncertainty about the estimated spectral signatures increases as the data quality decreases. In particular, when the per-pixel photon counts and the number of pixel per class decreases, it becomes more difficult to accurately estimate the $K$ spectral signatures, which in turn makes the classification less accurate.

![Image](image_url)

Fig. 7. Estimated spectral signatures of the classes associated with the materials #4 (red) and #9 (blue) of the Target 1 ($190 \times 190$ pixel scan). This figure compares the estimated spectral signatures using 8, 16 and 33 wavelengths and with an average of 1 detected photon per pixel and per band. The vertical dashed lines represent the selected wavelengths when considering only 8 spectral bands. The solid lines represent the posterior mean while the coloured regions represent confidence intervals of $\pm 3$ standard deviations around the posterior mean.
Figure 7 shows that reducing the number of spectral bands, and thus the spectral resolution, also reduces our ability to discriminate different material. For instance, if we only consider half of the spectral bands used in right-hand side sub-plot of Fig. 7, i.e., if we use only four wavelengths and do not consider wavelengths around 540 nm, the observable spectral signatures of materials #4 and #9 become extremely similar, which, when combined with extremely sparse photon counts, explains why the corresponding classes are merged. It is important to note that the confidence intervals depicted in Figs. 6 and 7 are obtained directly from the algorithm output, without additional cost. Such extra information would not be easily accessible using standard optimization techniques.

![Figure 7](image)

**Fig. 8.** Analysis of Target 2 shown in Fig. 2(b). The first column depicts the estimated depth profile (in mm), the reference range being arbitrarily set to the backplane range. The second depicts color classification maps. As in Fig. 4, the third column depicts the spectral signatures of the most prominent classes, projected onto the first and second axes obtained using PCA. Each of these subplots illustrates the similarity between the estimated spectral signatures. Rows a) and b) depict results obtained with an average of 1 detected photon per pixel, for each spectral band.

Figure 8 shows depth estimation and spectral classification results obtained using Target 2, which was composed of clay objects with more distinct colors than those used in Target 1. In addition to the higher number of spectral features contained in the second target, it is worth noticing that several objects have similar spectral signatures (different shades of green and brown) within the wavelength range considered (i.e. 500 - 820 nm) (see third column of Fig. 8). Moreover, other objects produce low reflectivities within the wavelength range considered (the blue and violet objects have higher reflectivities at wavelengths below 500 nm). In contrast to the results depicted in Fig. 4, these results have been obtained with an average of 1 detected photon per pixel, for each spectral band. As the number of detected photons and the number of bands decrease, spectrally similar classes are merged. However, even with only 4 bands (Fig. 8(b)), the proposed method is still able to discriminate materials which are difficult to distinguish visually.

For completeness, Fig. 9 presents the classification results obtained for the second target with...
33 and 4 spectral bands. Here, the classification problem becomes even more challenging due to the higher number of spectrally different objects and the high similarity between several spectral signatures (see Fig. 10 which shows the estimated spectral signatures of the most representative classes associated with the Target 2). Indeed, although three materials (#1, #2 and #14 in Fig. 2(b)) present significantly different signatures, the other objects present stronger similarities and are also less reflective within the spectral range used here.

Fig. 9. Example of spectral classification of Target 2 shown in Fig. 2 (b) with an average of 10 and 1 detected photons per pixel, for each spectral band.

Fig. 10. Estimated spectral signatures of the most representative classes associated with the Target 2, with an average of 10 detected photons per pixel, for each spectral band and using 33 wavelengths. The line colours correspond to the colours used in the classification maps in Fig. 9 and the numbering of the legend corresponds to the numbering using in Fig. 2 (b)
4.2. **Depth estimation**

In this section, we illustrate the benefits of the proposed methods for enhanced range profile analysis. Figs. 11 and 12 illustrate the visual degradation of the range estimation as the overall number of detected photons decreases, either because of the reduction of the number of bands, the per-pixel acquisition time or the number of observed pixels.

### Fig. 11. Estimated range profile of Target 1 shown in Fig. 2(a) with an average of 10 and 1 detected photons per pixel, for each spectral band. As depicted in Table 1, ranging performance degrades when reducing the number of spectral bands from 33 to 4. Dark (resp. white) pixels correspond to pixels the closest (resp. the farthest) from the imaging system. Note the depth scale is in mm.

<table>
<thead>
<tr>
<th>Spectral Bands</th>
<th>10 photons</th>
<th>1 photon</th>
</tr>
</thead>
<tbody>
<tr>
<td>33 wavelengths</td>
<td>190 x 190 pixels</td>
<td>190 x 190 pixels</td>
</tr>
<tr>
<td>4 wavelengths</td>
<td>48 x 48 pixels</td>
<td>48 x 48 pixels</td>
</tr>
</tbody>
</table>

### Fig. 12. Estimated range profile of Target 2 shown in Fig. 2(b) with an average of 10 and 1 detected photons per pixel, for each spectral band.

The depth estimation performance is assessed quantitatively using the range root mean square...
errors (RMSEs) defined by

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (d_n - \hat{d}_n)^2},$$  \hspace{1cm} (4)$$

where $d_n$ (resp. $\hat{d}_n$) is the actual (resp. estimated) range associated with the $n$th pixel, for the different scenarios discussed above. The RMSEs obtained for the two targets are summarised in Table 1. The first values in each cell of Table 1 corresponds to the RMSE obtained with the proposed method, using a TV-based depth regularisation, while the second value is obtained by considering independent target ranges. The reference range profile has been obtained using multispectral Lidar measurements acquired over a long acquisition time (33 wavelengths, with an average of more than 1000 detected photons per pixel for each spectral band). This table shows that by considering spatially correlated depth parameter (the amount of correlation being automatically adjusted by the proposed algorithm) yield a significant improvement of the RMSE, in particular when the data quality decreases. When the number of detected photons increases, the benefits of the regularization reduces. These results show that the RMSEs remain below 3 mm when more than 8 photons are detected per pixel (whatever the corresponding wavelengths). When the data sparsity increases further, it becomes more challenging to accurately recover the range profile. Moreover, this table shows that considering multiple wavelengths while keeping the same overall number of detected photons allows more robust range estimation. For instance, the depth RMSEs are smaller with 33 detected photons spread across 33 wavelengths than with 40 photons spread across 4 wavelengths (see 190 $\times$ 190 pixels scans in Table 1).

Table 1. Range root mean squared errors (RMSEs, in mm) for both targets, with different number of wavelengths, pixel format and photon number per pixel. The first (resp. second) numbers stand for the RMSEs obtained with (resp. without) the TV-based range regularization.

<table>
<thead>
<tr>
<th>Object</th>
<th>No. of bands</th>
<th>10 photons</th>
<th>1 photon</th>
<th>1 photon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>pixels</td>
<td>pixels</td>
<td>48 x 48</td>
</tr>
<tr>
<td>Target 1</td>
<td>33</td>
<td>0.35/0.55</td>
<td>0.74/1.75</td>
<td>0.86/1.59</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.46/0.77</td>
<td>1.10/3.60</td>
<td>1.17/2.90</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.68/1.14</td>
<td>2.25/8.84</td>
<td>1.75/8.39</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.03/1.74</td>
<td>6.02/21.28</td>
<td>5.42/21.40</td>
</tr>
<tr>
<td>Target 2</td>
<td>33</td>
<td>0.49/0.65</td>
<td>0.92/1.91</td>
<td>1.11/1.82</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.65/0.92</td>
<td>1.31/3.95</td>
<td>1.58/3.96</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.97/1.41</td>
<td>2.67/9.75</td>
<td>2.34/9.34</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.61/2.35</td>
<td>8.06/20.78</td>
<td>6.70/21.00</td>
</tr>
</tbody>
</table>

In a similar fashion to the spectral analysis, the proposed method can be used to asses the range estimation performance on a pixel by pixel basis. In particular, the generated samples can be used to provide a posteriori measures of uncertainty about each range parameter. For instance, Fig. 13 compares range marginal a posteriori probability density functions (pdfs) for a central pixel of the material #2 of the Target 2. This figure confirms that that increasing the number of detected photons generally reduces estimation errors. Increasing the number of spectral bands generally reduces estimation uncertainty (distribution centred closer to the true value) and yield more concentrated distributions. Moreover, for a given number of wavelengths, increasing the per-pixel acquisition time further reduces uncertainty. Note that the distributions in Fig. 13 are obtained from a single observation, conversely to the distribution of the marginal MAP depth estimator which would be obtained by analysing the position of the mode of the distributions in
Fig. 13 for many noise realizations. Thus, the potential error between the mode of the distribution and the true value of the depth parameter is mainly due to the observation noise and the bias induced by the TV-based regularization.

![Fig. 13. Example of range marginal a posteriori pdfs as a function of the number of spectral bands considered, with an average of 10 (left) and 1 (right) detected photons per pixel, for each spectral band (190 × 190 pixels scan). The dashed lines represent the actual position of the target with respect to the backplane (which corresponds to \( d_n = 0 \)).](image)

5. Conclusion

In this paper, we demonstrated the spectral classification of complex scenes from single-photon data constructed from less than one photon per pixel per spectral band. Accounting for the spatial organization of the pixels via Markov random fields allowed us to discriminate different natural and man-made materials that were spectrally similar, using only four separate wavelengths. We proposed a method capable of estimating range profiles in addition to classification maps. We also showed that spreading the energy across multiple wavelengths generally increases the probability of object detection (by increasing the probability of using wavelengths for which objects have a high reflectivity) and thus leads to more robust range estimation. Future work includes the development of efficient computational methods (e.g., by generalizing the proposed method) for applications where the ambient illumination and detector dark counts cannot be neglected, such as outdoor long-range imaging applications, for example remote mapping of foliage and agricultural regions. Accounting for reflectivity distortion in the low photon count regime, e.g., due to photon scattering through obscurants (fog/water) and different object orientations is also an interesting problem that deserves further attention. Finally, it is worth noting that although the proposed method performs simultaneously the color classification and the range estimation, the color classification does not rely on the knowledge of the depth profile. Thus, the proposed method can be applied to data which do not contain timing information and can have a significant impact in other imaging applications where single-photon images are used, for example for analysis of dynamic scenes (where the low spatio-temporal light field can be segmented into piece-wise constant regions) or low-light level microscopy. The method can also be extended to other types of particle-based (e.g., neutron spectroscopy) imaging and sensing strategies.
Funding

UK Engineering and Physical Sciences Research Council (Projects EP/K015338/1, EP/J015180/1, EP/M01326X/1, and EP/N003446/1); UK Royal Academy of Engineering (Research Fellowship scheme RF201617/16/31).