Spatial modelling of multi-layered LiDAR images using reversible jump MCMC

S. Hernandez-Marin[†], A.M. Wallace[†], G.J. Gibson[‡]
[†]Joint Research Institute on Signal and Image Processing, School of EPS
[‡]Maxwell Institute for Mathematical Sciences, School of MACS
Heriot-Watt University, Edinburgh EH14 4AS
{snh3,a.m.wallace,g.j.gibson}@hw.ac.uk

Abstract

3D imaging LiDAR systems have the potential to acquire multi-layered 3D image data; that is rather than store a single depth value at each pixel, it is possible to store the range to more than one surface within the pixel view direction. Multiple returns are possible at a single pixel when imaging through transparent surfaces, for example when acquiring depth images of cars or buildings that have windows, in which case it is possible to record both external and internal structure. Multiple returns are also possible when the pixel field of view encompasses more than one opaque surface. However, to build such multi-layered 3D images, we need to think of new ways of processing the LiDAR data.

In this paper, we present a unified theory of pixel processing for such data. This is based on a reversible jump Markov chain Monte Carlo (RJMCMC) methodology extended to include spatial constraints by a Markov Random Field with a Potts prior model. We consider two distinct proposal distributions, based on spatial mode jumping and spatial birth/death processes respectively. We also include a delayed-rejection step in the RJMCMC algorithm to improve the estimates of the range and reflectance of each surface element. Our methodology is demonstrated on both photon count and burst illumination LiDAR data.

1 Introduction

Our intention is to characterise completely all the 3D surfaces viewed by a Light Detection and Ranging (LiDAR) system. LiDAR works on the principle of time-of-flight, that is the range to the surface can be computed by measuring the go-return time of a laser signal when it impinges on a surface in the field of view (FOV). However, there is no reason why we cannot measure range to more than one surface along the same pixel field of view, if there are transparent surfaces or simply more than one surface in that FOV. Moreover, most scenes contain spatial patterns that have strong dependencies between different pixels. Some pixel configurations are more likely than others since multiple returns can be concentrated in certain regions and completely absent in others. There are many reasons why this occurs, for example, when mapping the pixels to landscape patterns, if a pixel is identified as 'water', it will be most likely surrounded by the same class of pixel. Further, different parts of an object

are related through geometric constraints. If these spatial interactions can be modelled, the classification accuracy can be improved [11].

Many spatial problems are inherently multivariate, in that more than one variable is measured at each spatial location. Multivariate data analysis allows users to display many different spatial data layers. Mardia [8] introduced a multivariate Markov random field (MRF) model for image processing although this work received little attention, due primarily to computational difficulties. Current interest in the analysis of multivariate lattice data has been concentrated on remotely sensed data, especially multispectral images [9], for target detection, scene classification and segmentation.

Often, this multivariate data arises from a time series of measurements. Burst Illumination Laser (BIL) [1] and Time Correlated Single Photon Counting (TCSPC) [7] are examples of 3D LiDAR techniques that can acquire both depth and reflectance images of objects. Fig. 1 shows an example of a single frame acquired with a BIL system and a histogram of integrated intensities from one pixel of Fig. 1(a). As the timing of the camera shutter is varied on the returned pulse, so the intensity rises and falls depending on the depth of the observed surfaces. Fig. 2 shows a small section of a TCSPC pixel image, in which the multiple returns at each pixel are clearly visible as distinct peaks in the photon histograms. Briefly, each pixel records a multivariate measurement which can be considered as an observed photon or intensity histogram which in turn is considered as a sample of a non-normalized statistical mixture distribution. The returns present come not only from the first surface encountered by the projected laser signal, but also from subsequent surfaces in its path. A simple approach to the problem would be to treat the time series of observations at a given image pixel as independent and identically distributed to that for all other pixels, and to analyse them with an appropriate model to understand the underlying theory of the data points. To model such data one can use a parametric approach [7] making use of appropriate mixture distributions and then use the techniques described therein to obtain estimates of the different parameters. However, such an approach can give suboptimal results because it neglects the correlations between the parameters in neighbouring pixels. The contribution of this paper is to incorporate such spatial constraints in the context of RJMCMC pixel processing.
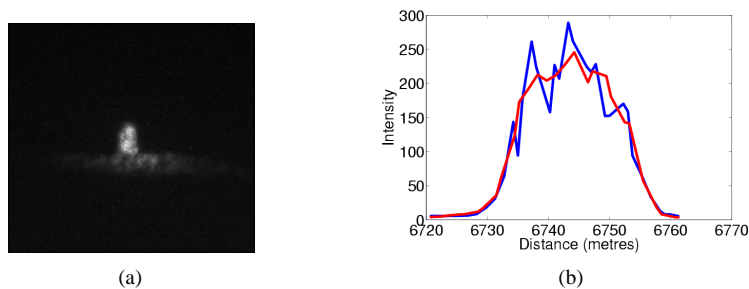

<div align="center">(a)   (b)</div>

Figure 1: (a) Single frame BIL image of a trig. point at distance of 6.6km (b) Variation of pixel intensity on the trig point as a function of distance from a BIL system (blue) and final fit from RJMCMC estimation (red)

Spatial dependencies have been introduced in a mixture distribution to take into account spatial heterogeneity using a Bayesian approach [2, 5]. However, the assumption is that the different observations come from a unique mixture distribution in which the number of com-

ponents and some of the parameters (maybe all) are variable. In these approaches, the spatial dependencies were introduced either through the weights of the mixture distribution [2] or through the different allocation variables [5]. Posterior inference was performed using RJM-CMC algorithms. Initially, one may consider applying a multivariate MRF to model the spatial dependencies among the different parameters of the different mixture distributions. However, it is not clear how to use these structures in a variable-dimension setting where the different peaks are continuously created or deleted to better explore the space of solutions.

Hence, we consider a complex scenario where spatial interaction is present. To analyse TCSCP and BIL data, we incorporate spatial constraints in a multi-layered image for the first time. Each pixel is a multivariate record, so we use a mixture distribution for each pixel instead of a "global mixture" [5] that includes all the possible returns. To explore the space of solutions, we apply to single pixels a Bayesian statistical approach based on RJMCMC techniques [10] to assess the number, positions and amplitudes of the returned signals from target surfaces [7]. We also include spatial dependencies between pixels that are based on the number of elements of the mixture through the prior distribution as in [6]. However, the major contribution of this paper is to incorporate two new moves within the RJMCMC algorithm, "spatial mode jumping" and a spatial birth/death process that incorporates spatial information in what is now a mixture of proposal distributions. We also introduce a delayed-rejection step [4] for variable-dimension setting to allow a kind of learning process.

2 Model

Consider a rectangular grid of pixels, labeled $m = 1, 2, \ldots, N$. At each pixel, a multivariate measurement $y^m = (y_1, y_2, \ldots, y_t)$ is available. Each multivariate measurement is a histogram of photon counts or intensities in which y_i^m is the value recorded in channel i, $i = 1, \ldots, t$. The exact temporal form of the photon count and the intensity histograms are unknown, as they depend respectively on the detector response and camera shutter (time-gate). To interpret such data we follow initially the same approach as [7]. For the TCSPC data, we have used the parametric form of the expected temporal variation of the photon count distribution, employing four piecewise exponential functions. In this study, we assume that the shape parameters of the returned pulses are fixed and known from an instrumental response. For BIL data, we use a look-up table of an instrumental function to interpret the histogram of intensities. This is acquired from the response of the BIL system to a Lambertian reflecting surface at a similar range to the object of interest.

Since our methodology is general, and can be applied to any pulsed LiDAR system, we use the notation f_{op} when referring to the LiDAR operating model. Furthermore, several signals can be present in the same histogram and these will be observed against a finite background level, B_m, whose expected value is considered as a constant across all the channels. The observed histogram in pixel m, $\mathscr{F}_m(i, k_m, \beta_m, t_{0_m}, B_m)$, can be considered as a sample of a statistical mixture distribution with density

$$\mathscr{F}_m(i, k_m, \beta_m, t_{0_m}, B_m) = \sum_{j=1}^{k_m} \beta_{mj} \cdot f_{op}(i, t_{0_{mj}}) + B_m \tag{1}$$

where k_m is the number of peaks, $\beta_m = (\beta_{m1}, \beta_{m2}, \ldots, \beta_{mk_m})$ is a vector of amplitude factors and $t_{0_m} = (t_{0_{m1}}, t_{0_{m2}}, \ldots, t_{0_{mk_m}})$ is a vector of the time of the peak maxima.

If the time resolution is sufficiently fine, the recorded value y_i^m can be considered to be a random sample of a Poisson distribution with intensity equal to $\mathscr{F}_m(i, k_m, \beta_m, t_{0_m}, B_m)$, which depends on the different model parameters.

$$y_i^m \sim Poisson(\mathscr{F}_m(i, k_m, \beta_m, t_{0_m}, B_m)) \tag{2}$$

To construct a likelihood function, we make two further assumptions. First, the observations in each channel i of the histogram are conditionally independent given the value of the parameters. Second, the spatial dependencies are included in Bayes' equation through parameters representing spatial interactions. Hence, the different $Y = (y^1, y^2, \ldots, y^N)$, are spatially independent given the parameters. The likelihood for the total array is expressed by the following equation

$$L(Y|k, \Omega, T, B) = \prod_{m=1}^{N} \prod_{i=1}^{i_{\max}} e^{-\mathscr{F}_m(i, k_m, \beta_m, t_{0_m}, B_m)} \frac{\mathscr{F}_m(i, k_m, \beta_m, t_{0_m}, B_m)^{y_i^m}}{y_i^m!} \tag{3}$$

where $k = (k_1, \ldots, k_N)$, $\Omega = (\beta_1, \ldots, \beta_N)$, $T = (t_{0_1}, \ldots, t_{0_N})$ and $B = (B_1, \ldots, B_N)$.

3   Bayesian Inference

The objective is inference about the parameters of Eq. 3 to obtain accurate estimates of the number of peaks, position, amplitude and background of the returned signals. If we do not have any spatial contextual information then these unknowns are regarded as drawn from independent prior distributions with full joint prior distribution

$$\pi(k_m, \beta, t_0, B) = \left(\frac{1}{k_{max}}\right) \left(\frac{1}{i_{max}}\right)^k f_G(B_m|c, d) \prod_{j=1}^{k} f_G(\beta_{mj}|a, b) \tag{4}$$

where f_G is the probability density function (pdf) of a gamma distribution.

We consider that observations that correspond to nearby locations are more likely to have similar numbers of peaks than observations from locations that are far apart. Therefore, we generalise the prior distribution given by Eq. 4 including a penalty function, the Potts model, that discourages adjacent pixels from having different numbers of peaks. The Potts model is a generalisation of the Ising model explained in [5] in which the random variable is allowed to have more than two different values. The Potts model has been used previously in image processing applications and in disease mapping applications to model allocation variables. We follow the formulation [5] in which k is modeled jointly

$$p_1(k|\psi) \propto e^{\psi \cdot U(k)} \tag{5}$$

where $U(k) = \sum_{m \sim m'} I[k_m = k_{m'}]$ are the number of like-labeled neighbouring pairs in the configuration $k = (k_1, k_2, \ldots, k_N)$. The parameter ψ is nonnegative and controls the amount of "smoothing", that is, $\psi = 0$ corresponds to a priori spatial independence of the number of peaks and as $\psi \to \infty$ we favour patterns where the number of peaks in neighbouring pixels tend to be similar.

The full joint prior distribution incorporating spatial constraints can be modeled as

$$f(k, \Omega, T, B) = p_1(k|\psi) \times \prod_{m=1}^{N} \pi(k_m, \beta_m, t_{0_m}, B_m) \tag{6}$$

Using equations 3, 6 and Bayes' theorem the target distribution can be expressed as

$$\pi(k,\Omega,T,B|Y) \propto L(Y|k,\Omega,T,B)f(k,\Omega,T,B) \tag{7}$$

since we just have to know the posterior distribution up to a normalizing constant.

Simulation from the joint probability distribution is difficult. Therefore, we update the different parameters using their respective full conditional distributions. We define a Markov random field with a second-order neighbourhood. The Hammersley-Clifford theorem [3] ensures that we can use the full conditional distribution of Eq. 5.

$$p_m(k_m|\ldots) \propto \exp(\psi \sum_{m' \in \delta_m} I[k_m = k_{m'}]) \tag{8}$$

where δ_m denote the neighbours of m. In this way, a rather complex multivariate probability distribution of a MRF can be obtained by successive simulations from the full conditional distributions.

4 Data Analysis

The Bayesian models used in this work are too complex to be amenable to analytical calculations. RJMCMC techniques [5] allow us to infer the number, positions and amplitudes of the returned signals from target surfaces. We allow moves between state spaces with different dimensionality, which in our case corresponds to changing the number of viewed surfaces. When the dimension is fixed, parameter moves improve the parameter estimates, corresponding to surface range (time of arrival), reflectance (amplitude) and background level. RJMCMC is our preferred method for updating beliefs in response to new information and incorporates prior knowledge. RJMCMC techniques allow us to explore the full posterior distribution of the parameters of the mixture distribution of Eq. 1, given the data values Y supplied by the different LiDAR histograms. The parameters are estimated from the values of a Markov chain whose limiting distribution is a target distribution π. π arises from the posterior distribution defined by Eq. 7. We use simpler versions of π involving only full conditional distributions. The Markov chain constructed involves moves of various types. These are: (a) updates to parameters Ω, T, B, (b) the random birth of a peak, (c) death of a peak, (d) the random splitting of a peak into two peaks and (e) merging of two peaks into a single peak. Movements (b), (c), (d) and (e) are governed by a probability explained in [10]. The implementation of these moves is described in [7].

Our approach also includes a delayed-rejection stage in the RJMCMC algorithm as proposed in [4] in order to improve the mixing of the Markov chain, improving the estimates of the different parameters. In the delayed-rejection strategy, if a candidate move is rejected, we make another attempt to move using a second proposal instead of turning to the next transition. There is no restriction on the number of stages used, but we use only two stages to reduce the computational load. Using delayed-rejection, three different acceptance probabilities have to be calculated. Using Green and Mira's notation [4] for general state spaces, to move from a state x to a state xx, we propose to draw a random number u_1 from a known density g_1. The new stage xx is calculated as $xx = h_1^+(x,u_1)$. The reversed move, from xx to x, is performed by drawing u_1' from g_1' and calculating $x = h_1^-(xx,u_1')$ where h_1^+ and h_1^- are deterministic mappings. The acceptance probability of this first stage is given by

$$\alpha_1(x,xx) = min\left\{1, \frac{\pi(xx)g_1'(u_1')}{\pi(x)g_1(u_1)}\left|\frac{\delta(xx,u_1')}{\delta(x,u_1)}\right|\right\} \tag{9}$$

If this candidate move is rejected, a new move to state z is proposed in a similar way to the previous candidate move; this move is accepted or rejected with a probability given by

$$\alpha_2(x,z) = min\left\{1, \frac{\pi(z)\tilde{g}_1(\tilde{u}_1)\tilde{g}_2(\tilde{u}_2)\{1-\alpha_1(z,xx^*)\}}{\pi(x)g_1(u_1)g_2(u_2)\{1-\alpha_1(x,xx)\}}\left|\frac{\delta(z,\tilde{u}_1,\tilde{u}_2)}{\delta(x,u_1,u_2)}\right|\right\} \tag{10}$$

where $z = h_2^+(x,u_1,u_2)$ and $x = h_2^-(z,\tilde{u}_1,\tilde{u}_2)$. The values $\tilde{u}_1,\tilde{u}_2$ and u_2 are drawn from $\tilde{g}_1,\tilde{g}_2$ and g_2 respectively.

To obtain the numerator of the second acceptance probability we need to calculate a third acceptance probability, $\alpha_1(z,xx^*)$, corresponding to a fictional stationary Markov chain started in z which proposes a move to xx at the first stage, rejects it and accepts a second stage move to x. Although this chain is not really implemented, it is needed to ensure the reversibility condition and therefore preserve the stationary distribution. The expression of the acceptance probability of the virtual stage is identical to that of Eq. 9 with an appropriate change of variables.

In practice, in a 3D secene, objects that have the same attributes tend to cluster in space. Hence, groups of pixels with the same number of peaks and similar parameter values are expected to occur together. Spatial interactions for positions are also expected to occur. However, only spatial interactions for the number of peaks are represented in the prior. Furthermore, the Potts model constrains the number of peaks without considering the relative positions and amplitudes of the current pixel with respect to its neighbouring pixels. Since such contextual information is not available in the prior, we may force situations in which the state space is not explored properly. Therefore, we suggest a proposal distribution for position moves which can exploit the fact that we expect data to show correlated positions. Thus, we have incorporated two new moves, called "spatial mode jumping" and a spatial birth/death process, within the RJMCMC algorithm described in [7]. The "spatial mode jumping" move proposes an update to the position of a peak in pixel m so that its position corresponds to a random perturbation of the position of a peak in pixel m' with some probability. The spatial birth/death move penalises the creation or removal of a peak given the information of the neighbourhood of the pixel. In the spatial birth process, the proposed values for the amplitude and the position of the new peak are drawn from two mixture distributions whose elements incorporate spatial information from neighbouring pixels in a similar way to that of "spatial mode jumping". For a death, one of the current peaks is chosen at random to be removed. Furthermore, the birth and death moves incorporate a delayed-rejection step which learns from the previous rejected value and therefore allows some kind of learning. To model these moves, we use a mixture of proposal distributions which takes account of the current values of the actual peak m and the second-order neighbouring pixels (to reduce algorithm complexity) given by the following expression

$$Q(\cdot) = w_m \cdot q_m^1(\cdot) + \sum_{l \in \delta_m} w_l \left(\sum_{t=1}^{k_{max}} v_{lt} \cdot q_t^2(\cdot)\right) \tag{11}$$

where q^1 and q^2 are proposal distributions whose expressions depend on the move we are performing, k_{max} is the maximum number of peaks and w and v are weights that satisfy $\sum_{l=1}^{9} w_l = 1$ and $\sum_{t=1}^{k_{max}} v_{lt} = 1$ respectively. These weights define positive probabilities in such a way that at each step one of the proposals is selected according to these probabilities. The first term of Eq. 11 ensures that a standard independent pixel move is going to be proposed when no peaks are available in the neighbourhing pixels. The second term includes spatial information in the proposal distribution.
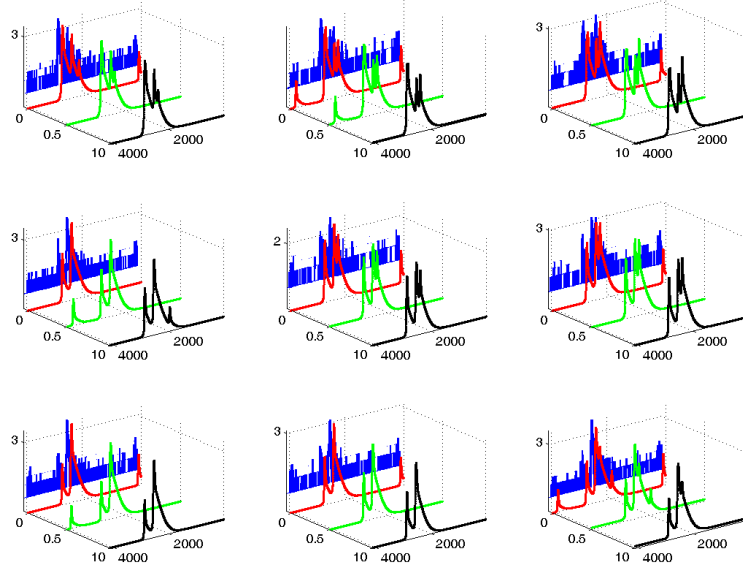
5 Experimental Results

Figure 2: Subimage array of photon count histograms (in blue) of the structure of Fig. 4(a) and final fit without spatial constraints $\psi = 0$ (red), with $\psi = 0.5$ (green) and $\psi = 10$ (black). The vertical axis represent photon counts (in log scale for convenience) whereas the horizontal left axis represents different smoothing parameter values, ψ, and the horizontal right axis represents temporal channels.

In this experiment, we analyse TCSPC and BIL images that correspond to a pixel array of 50 by 50 histograms of photon counts and reflectance/intensity values respectively. The results correspond to 1000 sweeps of the algorithm described in section 4. We infer the number of peaks as that corresponding to the highest marginal posterior probability, $\hat{k} = argmax \, p(k|y)$. Once we determine the number of peaks, we extract estimates of the parameters from $p(\phi|k,y)$ by setting $p(\phi|k = \hat{k},y)$. The values of the parameters are estimated as the mean values of the samples that correspond to such a parameter subspace. The TCSPC histograms come from the structure shown in Fig. 4(a). This data has been analysed with the following values for the parameter $\psi = \{0.5, 5, 10\}$ of the Potts model defined by Eq. 5. The maximum number of peaks in this experiment was set to 10 and the initial number of peaks was selected randomly. Fig. 2 shows an example of subimages of 3 by 3 pixels acquired by TCSPC as well as the final fit obtained with and without spatial constraints. Fig. 3 3(a) and 3(b) display the estimated number when no spatial constraints are considered of peaks and when the smoothing parameter is equal to $\psi = 10$ respectively. As can be seen when using a Potts model the algorithm tends to form clusters, that is, areas with the same number of estimated peaks are connected by paths from neighbour to neighbour.

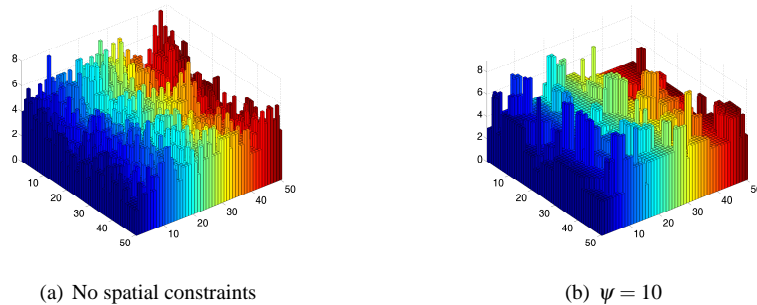(a) No spatial constraints   (b) $\psi = 10$

Figure 3: Estimated number of peaks of the TCSPC data using RJMCMC

Fig. 4(c) and 4(d) show three layer representations of a scene in which a porcelain cat is placed inside a greenhouse structure. The surface data are represented as point clouds without and with spatial constraints [1] $\psi = 10$. The different surfaces are difficult to perceive with this representation due to the display facilities of Matlab, which was used for code development. However, the shape of the toy cat is discernible and is shown without the other layers in Fig. 4(b). Initially one may think that the number of maximum returns is restricted to 2. Nevertheless, one pixel may have up to 9 returns. The reason of this is due to two main factors. First, our system presents multiple path reflections due to arrangement of the experiment. Second, the beam is impinging on a target with surfaces distributed in depth. Increasing the smoothing parameter enforces corresponding number of surface returns between adjacent pixels; increasing this value further means that the posterior distribution is dominated by the prior information. On the other hand, "false" returns will not be removed if the smoothing parameter is very small. As can be seen from Fig. 4(c) and 4(d), the spatial constraints used "kill" those returns which do not follow any spatial pattern.

The performance of the model is also illustrated on BIL data. The data set chosen is that corresponding to a UK Ordinance Triangulation Point (trig point), imaged at a distance of approximately 6.6km from the sensor, shown in Fig. 4(e). As such the LiDAR data is very different from that acquired in a laboratory setting, as the laser is subject to sever turbulence effects. This data presents only one unique return because the surfaces are opaque.. Therefore, only one peak and background returns were considered. The spatial information was incorporated through the proposal distributions and small values of the smoothing parameter $\psi = \{0, 0.5\}$ are considered. The $\psi = \{0.5\}$ value helps to "smooth" the number of peaks in several places where previously it was considered there was only background. If the smoothing parameter is increased, the results obtained are similar to those obtained with $\psi = 0.5$. Fig. 4(f) shows a depth reconstruction of the trig. point.

6 Conclusions

We have described the development and application of RJMCMC techniques incorporating spatial contextual information to process time-of-flight LiDAR data. This spatial information is incorporated on the prior distribution through a Markov Random Field on the number of

---

[1] Only one spatial-constrained multilayered image is shown due to paper length limitations
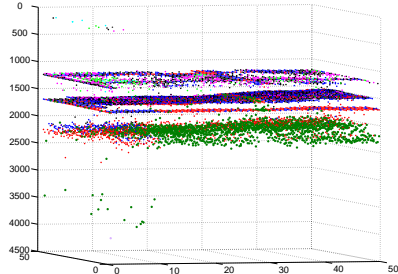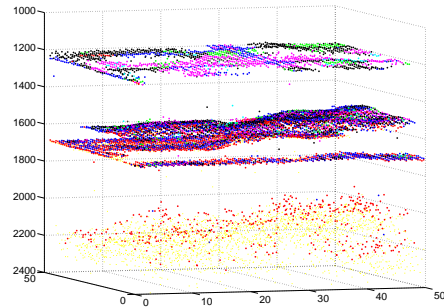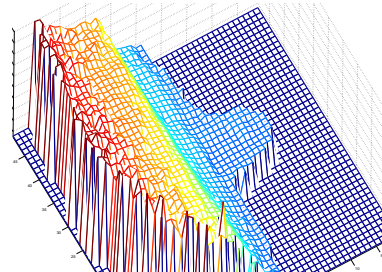
(a)
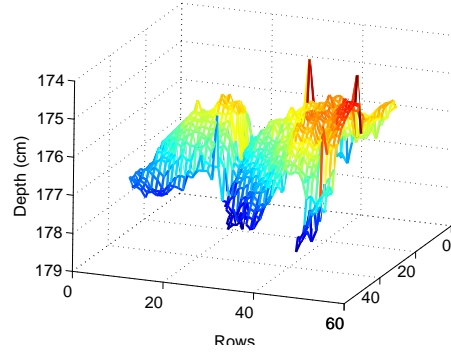
(b)

(c)

(d)

(e)

(f)

Figure 4: (a) Image of complex multiple return structure composed of a toy cat and a semi-transparent toy greenhouse (b) Mesh of the toy cat of (a) obtained with a TCSPC system (c) and (d) Multiple-layer structure of the experiment (a) with no spatial constraints and smoothing parameter $\psi = 20$ (e) Details from a trig point (distance ~ 6.6km) (f) Depth image obtained with a BIL system using RJMCMC with spatial constraints

peaks, and on the proposal distribution through the development of two new moves: the spatial mode jumping and the spatial birth/death processes. This spatial information can be

used to reduce the influence of clutter of unknown origin by eliminating possible false returns. Further, we have described a delayed-rejection stage which allows us to perform rudimentary learning and therefore improve the performance of this stage.

Our techniques are generally applicable, and have been demonstrated on LiDAR data using both single photon counting and variable temporal gating to extract range measurements. The results are excellent, and show that it is possible to resolve multiple returns and hence characterise objects distributed in 3D space, extracting the underlying spatial structure. For TCSPC LiDAR data, the imaging of multiple surfaces shows a capability well in advance of most 3D LiDAR systems. For BIL data, the detailed and accurate imaging of a the trig point concrete surface and hill grass and earth through a turbulent medium at several Km is also a significant achievement.

References

[1] P. Andersson. Long-range three-dimensional imaging using range-gated laser radar images. *Opt. Eng*, 45(3):1–10, 2006.

[2] C. Fernandez and P. J. Green. Modelling spatially correlated data via mixtures: a Bayesian approach. *J. Roy. Statist. Soc. Ser. B*, 64:805–826, 2002.

[3] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell*, 6:721–741, 1984.

[4] P. J. Green and A. Mira. Delayed rejection in reversible jump Metropolis-Hastings. *Biometrika*, 88:1035–1053, 2001.

[5] P. J. Green and S. Richardson. Hidden Markov models and disease mapping. *J. American Statistical Association*, 97:1055–1070, 2002.

[6] S. Hernandez-Marin, A.M. Wallace, and G.J. Gibson. Creating Multi-layered 3D Images Using Reversible Jump MCMC Algorithms. In *Advances in Visual Computing, LNCS, Springer-Verlag*, pages 405–416, 2006.

[7] S. Hernandez-Marin, A.M. Wallace, and G.J. Gibson. Bayesian analisis of Lidar signals with multiple returns. *IEEE Trans. Pattern Anal. Mach. Intell.*, in press, 2007.

[8] K. V. Mardia. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *J. Multivar. Anal.*, 24(2):265–284, 1988.

[9] H. Ren and C. I. Chang. A generalized orthogonal subspace projection approach to unsupervised multispectral image classification. *IEEE Trans. Geoscience and Remote Sensing*, 38(6):2515–2528, 2000.

[10] S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B*, 59:731–792, 1997.

[11] B. Tso and P. M. Mather. *Classification methods for remotely sensed data*. Taylor & Francis New York, 2001.