NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

A Model for Recognition of 3D Non-Dense Objects in Range Images

by

Oddvar Husby and Ulf Grenander

PREPRINT
STATISTICS NO. 15/2001
ISSN: 0804-9173

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2001/S15-2001.ps
Oddvar Husby has homepage: http://www.math.ntnu.no/~okh
E-mail: okh@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

A model for recognition of 3D non-dense objects in range images

Oddvar Husby
Department of Mathematical Sciences
NTNU, Norway

Ulf Grenander
Division of Applied Mathematics
Brown University, U.S.

Abstract

This paper discusses a deformable template approach to the problem of recognising three dimensional, non-dense objects in high-resolution laser range images. To model the infinite variability in object appearance we develop an imaging model based on a Poisson object process, assuming objects to consist of primitives distributed according to a non-homogeneous Poisson point process. We discuss some computational aspects of the model, and show how we can use the Metropolis-adjusted Langevin Algorithm (MALA) to generate samples from the posterior distribution. We show results applying the model to real laser range images of forest.

1 Introduction

Recognition of three-dimensional objects from remotely sensed scenes has received considerable attention over the last few years, especially in connection with automated target recognition (ATR). Laser radar imagery is particularly well suited for such tasks, as the need for modelling light-sources and reflectance is much less pronounced than is the case for intensity images. A laser range image is a collection of distances measured along rays emanating from the laser and indexed by azimuth and elevation angles,

$$r = \{r(\theta, \phi)\ ;\ \theta_1 \leq \theta \leq \theta_2,\ \phi_1 \leq \phi \leq \phi_2\}. \tag{1}$$

So far, much of the work has focused on detection based on non-contextual models (Green and Shapiro, 1994) and on classification based on the matching of range profiles (Koksal, Shapiro and Wells, 1999; Zhou, Liu and Wang, 2000; Webb, 2000; Nair and Aggarwal, 2000; Jacobs and O'Sullivan, 1997). The use of contextual models for more advanced tasks has been advocated by the group centred at Washington University, focusing on recognition of rigid objects in radar (Miller, Srivastava and Grenander, 1995; Srivastava, Miller and Grenander, 1997; Jacobs, O'Sullivan, Faisal and Snyder, 1997) and forward-looking infrared (Lanterman, Miller and Snyder, 1997) images.

We adopt the approach of the above-mentioned group, using the deformable template models introduced in Grenander, Chow and Keenan (1991) and Grenander (1993). These are highly structured probability models based on physical knowledge of the imaged objects, and containing significant contextual information. Objects are divided into classes taken from a finite alphabet, and for each class a priori knowledge on object shape is represented via a template I_0, a parametric model of a typical object. Natural variability in object shape is then modelled by applying a transformation group $\mathcal{S}$ to the template, generating a set $\{sI_0\ ;\ s \in \mathcal{S}\}$ of objects. Deformable template models have successfully been applied in two-dimensional applications, eg. for recognising hands (Grenander et al., 1991), locating and classifying cells (Rue and Hurn, 1999), and for estimating the outline of blood vessels (Hansen, Møller and Tøgersen, 2000; Husby, 2001). In 3D applications focus has been on estimating location and pose of rigid objects, ie. finding parameters in the special Euclidean group $\mathrm{SO}(3) \ltimes \mathbb{R}^3$ corresponding to rotations and translations of the template CAD surfaces representing the various objects present in the images, see eg. Srivastava (1996) for an overview.

However, lately there has been an increasing interest in modelling non-rigid objects and objects with a high degree of shape complexity (Jain and Dorai, 2000). This paper represents a step in that direction, since we are concerned with objects having highly variable shapes. This has implications both for the way objects need to be represented, and for the choice of transformation groups. The representation should be flexible enough to be able to capture the shape differences needed, while at the same time lend it self to efficient computations. Ideally, a representation should also possess certain invariances, such as being viewpoint independent. There exists a wealth of different representation (see Jain and Dorai (2000) for a review), and there is unlikely to be one representation being suitable for all object types and sensors. We follow a slightly non-standard path, letting a triangulated surface represent the main outline or envelope of the object, and modelling variation on a finer scale by assuming the object to consist of primitives placed at random positions within the region defined by the outline. We believe that this is an useful approach when dealing with objects that have a high degree of shape complexity or are otherwise non-standard, since it allows us to separate the two sources of shape variability, thus reducing the complexity of the recognition problem. We especially have in mind objects that are porous, having holes at random positions. The small scale variability due to the porosity will effect the performance of the recognition method, but is not of primary concern for us, as only the global shape features are relevant for comparison and classification of objects.

To our knowledge there has been little work on modelling of objects with high variability in density and appearance, although Larsen and Rudemo (1998) explicitly models the density of needles and branches in tree templates. However, when using range data it is sufficient to consider binary object models, since we assume the transmitted beam to be reflected fully when hitting the first object along its path. Thus we can assume an object to consist of geometric primitives placed at the positions of an inhomogeneous Poisson point process. This allows us to treat the small-scale variability pixel-wise instead of object-wise, inducing a probability density function π for the range at pixel i

$$\pi\left(r_i \mid \lambda\right) = \lambda\left(\theta_i, \phi_i, r_i\right) \exp\left(-\int_0^{r_i} \lambda\left(\theta_i, \phi_i, r\right) \mathrm{dr}\right). \tag{2}$$

The function λ is crucial, as it models the appearance of the objects. To reduce complexity and allow for efficient computations, we will assume the intensity function λ to be piecewise constant, with high intensity in regions occupied by the template objects.

We take a Bayesian approach, defining a posterior distribution on the transformation group $\mathcal{S}$, and using Markov chain Monte Carlo methods to explore it. In particular, we use the Langevin-Hastings algorithm suggested by Besag (1994). The algorithm is based on Langevin's stochastic differential equation (SDE)

$$dX(t) = \nabla H\left(X(t)\right) dt + \sqrt{2} dW(t), \tag{3}$$

which generates a Markov process with a stationary distribution proportional to $\exp(H(x))$, and which is shown to be superior to the Random-walk algorithm in many situations (Roberts and Tweedie, 1996; Christensen, Møller and Waagepetersen, 2000). However, additional complexity is introduced when the posterior distribution takes values on curved manifolds lacking the familiar vector space structures, as is often the case when the posterior is parametrised in terms of Lie groups. The theory for constructing SDEs on manifolds is well established (Kunita, 1984; Gliklikh, 1996), but only recently has it been put to use in a statistical setting for sampling Lie-group valued probability measures by means of Langevin's SDE (Piccioni and Scarlatti, 1994; Srivastava, Grenander, Jensen and Miller, 1999). These methods are described briefly in Section 4.

The range laser sensor is described in Section 3.2. Object representation using deformable templates is discussed in Section 2, while we present our point process imaging model in Section 3. In Section 5 we present some results using real range laser data of forest.

2 Object representation

We will use the global shape models developed from Ulf Grenanders pattern theory (Grenander, 1993) to analyse the imaged scenes. The concept of deformable templates presents an unifying way to analyse the variability on shape and occurrence of imaged objects. Let $\mathcal{A}$ be the finite set of possible object types. For each object type $\alpha \in \mathcal{A}$ we define a template $I_0(\alpha)$ constituting all object features affecting the imaging sensor, and representing a typical object in the class. For example, in target recognition the template can be a surface manifold representing the target shape. The variability in object shape within a class is accommodated by applying a transformation group $\mathcal{S}$ to the template. For each $s \in \mathcal{S}$, let $sI_0(\alpha)$ denote the action of s on the template; then the orbit

$$\mathcal{O}^\alpha = \{sI_0(\alpha)\ ;\ s \in \mathcal{S}\}$$

contains all possible object occurrences. We shall assume the group $\mathcal{S}$ to act transitively on $\mathcal{O}^\alpha$, so that each object occurrence can be uniquely represented by an element $s \in \mathcal{S}$.

When modelling objects with a high degree of shape complexity the transformation group needs to be very high-dimensional to be able to capture the variability. This has implications for the computational complexity of the recognition algorithm, while not necessarily leading to a higher classification accuracy. This is indeed the case in our example, where the objects have local shape features that affect the sensor, but that are otherwise of no interest. To get a compact and computationally efficient representation, we separate the two sources of variability and represent the global shape by a closed, piecewise smooth surface c_0 approximated via m triangular patches, each identified with a set of three vertices and a surface normal, see Fig. 1 for an example. By global shape we shall mean the features used for discriminating between different object types, and not features on a finer scale such as holes, irregularities at the boundary and so on. The surface encloses a region $G_0 \subset \mathbb{R}^3$, and the object itself is modelled as the collection of geometric primitives placed at the random positions of an Poisson process on G_0 as follows.

Let the region of interest be $W \subset \mathbb{R}^3$, and let Φ be an inhomogeneous Poisson process with intensity function $\lambda(x_1, x_2, x_3) = \lambda_{\mathrm{B}} \mathbb{I}_{W \setminus G_0} + \lambda_0 \mathbb{I}_{G_0}$, where $\mathbb{I}_A(x)$ is the indicator function having value one if $x \in A$, and zero otherwise. Furthermore, let $\{\chi_i\}_{i=1}^{n_W}$ be the collection of random positions of Φ in W, and associate with each point a geometric primitive P, inducing an Poisson object process

$$I = \bigcup_{i=1}^{n_W} P[\chi_i]. \tag{4}$$

Thus the template is a collection $I_0(\alpha) = (c_0, \lambda_0)$ of a closed surface c_0 representing the global shape of the object, and an intensity function λ_0 describing the density.

Next we introduce the transformation group $\mathcal{S}$ generating the object space. In our example, concerning recognition of trees, we have found it sufficient to consider simple, low-dimensional groups, letting $\mathcal{S} = \mathrm{A}(3) \times \mathrm{US}(1)$, the product of the affine group and the group of uniform scaling in one dimension. By a deformed template

$$sI_0(\alpha) = (s_1 \circ c_0, s_2\lambda_0)\,, \quad (s_1, s_2) \in \mathrm{A}(3) \times \mathrm{US}(1) \tag{5}$$

we shall mean the surface $s_1 \circ c_0$ formed by transforming all vertices and normals by s_1, together with the function $s_2\lambda_0$ obtained by scaling λ_0.

Taking a Bayesian approach we define conditional prior probability densities $\pi(s \,|\, \alpha)$ on the transformation groups, and a probability distribution $\pi(\alpha)$ on the set $\mathcal{A}$ of object classes. We choose the elements of the scale group $\mathrm{US}(1)$ to be Gamma distributed with hyperparameters p and q. The affine group

$$x \mapsto Ax + a, \quad A \in \mathrm{GL}(3), a \in \mathbb{R}^3 \tag{6}$$

represents translation, scaling, rotation, bending and skewing. We chose the elements to have independent, Gaussian distributed components, ie. $s \in \mathrm{A}(3)$ has density

$$\pi(s) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{12}\left(\frac{s_i - \mu_i}{\sigma_i}\right)^2\right) \tag{7}$$

with respect to Lebesgue measure. The means μ_i are chosen so that the distribution has the identity transformation as its mode.

3 Image formation and sensor modelling

When an object described by an deformed template $sI_0(\alpha)$ is mapped to an image r, two sources of variability is introduced. One is due to the description of the template as a random Poisson object process I, together with the projection of I onto an abstract true range image r^*. The second is measurement noise introduced when collecting the data, and is sensor dependent. We begin by describing the imaging formation model $\pi(r^* \,|\, s)$ and its properties in Section 3.1. The sensor likelihood $\pi(r \,|\, r^*)$ is described in Section 3.2.

3.1 Image formation

Throughout this section we will assume orthographic projection, although an extension to perspective projection is possible. Assuming the sensor to be far from the imaged objects, and the imaged region to be small, this assumption is reasonable. We identify the imaging plane of the sensor with the set $\Omega \subset \mathbb{R}^2$, and assume for simplicity continuously observed measurements. The ideal data is then $r^* = \{r^*(x_1, x_2) \; ; \; (x_1, x_2) \in \Omega\}$. Denote the axis orthogonal to the imaging plane by x_3, and consider a ray $\rho(x_1, x_2)$ emanating from the point (x_1, x_2) and going in the positive x_3-direction . The probability of observing a particular range $r^*(x_1, x_2)$ is then the probability of $r^*(x_1, x_2)$ being the site of the first event of the Poisson process Φ along the ray $\rho(x_1, x_2)$:

$$\pi_{x_1,x_2}\left(r^*(x_1,x_2) \,|\, s\right) = \lambda\left(x_1, x_2, r^*(x_1,x_2)\right) \exp\left(-\int_0^{r^*(x_1,x_2)} \lambda(x_1,x_2,x_3)\, dx_3\right). \tag{8}$$

By integrating over the imaging plane Ω we get the density for the whole image r^*:

$$\pi(r^* \,|\, s) = \exp\left(-\iint\limits_{\Omega}\int_0^{r^*(x_1,x_2)} \lambda(x_1,x_2,x_3)\, dx_3\, dx_2\, dx_1 + \iint\limits_{\Omega} \ln \lambda(x_1,x_2,r^*(x_1,x_2))\, dx_2\, dx_1\right). \tag{9}$$

This can be further simplified by assuming the density function to be piecewise constant. Then, letting $F = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \; ; \; (x_1, x_2) \in \Omega, \; 0 < x_3 \le r^*(x_1, x_2)\}$ denote the part of the world W that is observed under the sensor, and letting $\mathbb{R}^3 \supset G_s = s \circ c_0$ be the set occupied by the template, we define the intensity function as

$$\lambda(x_1,x_2,x_3) = \begin{cases} \lambda_{\mathrm{B}}, & (x_1,x_2,x_3) \in G_s^{\complement} \cap F \\ \lambda_0, & (x_1,x_2,x_3) \in G_s \cap F. \end{cases} \tag{10}$$

The density (9) then simplifies as follows.

Proposition 1
Let the density (9) have an intensity function as defined in (10). The density can then be written as

$$\pi\left(r^* \,|\, s\right) \propto \exp\Big(-(\lambda_0 - \lambda_B)\, m\left(G_s \cap F\right) + \ln\left(\lambda_0/\lambda_B\right)\, m\big(\mathcal{P}\left(G_s \cap \partial F\right)\big)\Big), \tag{11}$$

where $\partial F = \{(x_1, x_2, x_3) \in F \; ; \; x_3 = r^*(x_1, x_2)\}$, $\mathcal{P}$ *represents the orthographic projection* $(x_1, x_2, x_3) \mapsto (x_1, x_2)$, *and* $m(\cdot)$ *is Lebesgue measure.*

Evaluation of the density does in other word amount to calculating the volume and projected area of the deformed templates. This might be computationally expensive, especially calculating the area $\mathcal{P}(G_s \cap F)$ which may be highly irregular. Effort should be put into designing efficient ways of doing this.

The simulation algorithm described in Section 4 depends on gradient information in the posterior density. When the intensity function is as defined in (10), we can derive an explicit formula for a part of the gradient $\nabla \ln \pi(r^* \,|\, s)$. The result is as follows.

Proposition 2
Assume the deformed template shape $s \circ c_0$ *to be a closed, piecewise smooth surface with outward normal* n, *and let* G_s *be the region bounded by it. Denote by* $a_1, \ldots, a_{12}$ *a basis for the Lie algebra* $\mathfrak{a}(3)$ *of left-invariant vector fields on* $A(3)$. *Then*

$$\frac{\partial}{\partial s_k} \iiint\limits_{G_s \cap F} dx = \iint\limits_{\partial G_s \cap F} \langle n, a_k \circ x \rangle\, dx, \tag{12}$$

where $\langle \cdot, \cdot \rangle$ *is the usual inner product on* $\mathbb{R}^3$.

Proof. We sketch a geometric proof. Let $s_{\varepsilon,k}$ be equal to s, except for a small variation ε in the kth component, and let $G_{\varepsilon,k}$ be the volume enclosed by the corresponding deformed template. Then the gradient is proportional to the limit $\varepsilon \downarrow 0$ of the volume $(G_s \setminus G_{\varepsilon,k}) \cup (G_{\varepsilon,k} \setminus G_s)$. Represent the template surface c_0 by a triangulation $c_0 \approx \cup_{l=1}^{L} T_l$, with each triangle having area A_l, outward unit normal n_l and centre point p_l. The integral on the left hand side of (12) can then be approximated by the sum of parallelogram volumes $\sum_l \left\langle n_l, \frac{\partial s p_l}{\partial s_k} \right\rangle A_l = \sum_l \langle n_l, a_k \circ p_l \rangle A_l$, which in the limit $\varepsilon \downarrow 0$, $L \to \infty$ becomes the surface integral in (12). □

3.2 *Laser radar range images*

A coherent laser radar collects a range image by scanning a field of view, transmitting a single laser pulse for each pixel in a raster scan. The value at each pixel represents the time-of-flight between the peak of the transmitted pulse and the peak intensity of the video-detected intermediate frequency return waveform. The data used in this study were collected using a Riegl LMS-Z210 laser range-finder with a rotating mirror and an angular separation of 0.18 deg. The images contain 440×1440 pixels, and thus represents a field of view of $80°$ vertically and $259°$ horizontally. The operational range of the apparatus is approximately $2 - 200$ m. An image is represented as an array r defined on a lattice $\Lambda \subset \mathbb{Z}^2$.

Range images are subject to additive measurement noise and range fluctuations due to laser speckle (Goodman, 1975). The latter effect causes measurements far off the true range when a speckle fade occurs in conjunction with a noise peak (Green and Shapiro, 1992). To counter these effects we use the observation model of Green and Shapiro (1992), assuming the probability density of the measured range image r given the the true range image r^* to be a product of mixtures

$$\pi(r \mid r^*) = \prod_{i \in \Lambda} \left((1-p) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(r_i - r_i^*)^2 \right) + \frac{p}{R} \right), \tag{13}$$

where σ is a measure of the local range uncertainty, R is the operational range $R = r_{\max} - r_{\min}$, and p is the probability of an anomalous measurement. The parameters σ and p are system dependent, and approximate formulas are given in Shapiro, Reinhold and Park (1986).

4 Posterior sampling using the Langevin-Hastings algorithm

The a priori object model and the imaging models are combined to yield the a posterior distribution for the scene parameters s given the data r,

$$\pi(s \mid r) \propto \pi(r \mid r^*) \pi(r^* \mid s) \pi(s). \tag{14}$$

We will restrict attention to a single object, so s takes values in the product Lie group $\mathrm{A}(3) \times \mathrm{US}(1)$, while r is defined on the lattice Λ. The Langevin-Hastings algorithm presents a generic way of updating all parameters of a d-dimensional random vector S simultaneously, using gradient information to explore the parameter space. It is based on a discretization of the Langevin SDE which on vector spaces takes the form

$$dS(t) = \frac{1}{2} \nabla \ln \pi\left(S(t) \mid r\right) \mathrm{d}t + dW(t), \tag{15}$$

where $W(t)$ is a standard d-dimensional Wiener process. Assuming that the parameter vector S is defined on a vector space, and that the current state of the Markov chain is S_t, the algorithm proceeds by proposing a new state S' by going a distance along the gradient and adding Gaussian noise,

$$s' \sim \mathrm{N}\left(S_t + \frac{1}{2} h \nabla \log \pi(S_t \mid r), hI_d \right). \tag{16}$$

We denote the density of this proposal by $q(S_t, S')$. The new state S' is accepted with probability

$$\alpha(S_t, S') = \min\left\{ 1, \frac{\pi(S' \mid r)\, q(S', S_t)}{\pi(S_t \mid r)\, q(S_t, S')} \right\}, \tag{17}$$

otherwise the old state S_t is retained. The Markov chain then converges to the posterior distribution π, $\|\Pr(S_t \in \cdot | S_0 = s_0) - \pi(\cdot)\|_{\mathrm{TV}} \to 0$ for π-a.e. s_0.

Note: The affine group $\mathrm{A}(3) = \mathrm{GL}(3) \ltimes \mathbb{R}^3$ is strictly speaking not a vector space, it is not closed under addition. Translation on matrix Lie groups is defined by matrix multiplication, and the Langevin equation should be modified accordingly. In general, let G be a d-dimensional matrix Lie group, and let $E_1, \ldots, E_d$ be a basis for the Lie algebra $\mathfrak{g}$ of left-invariant vector fields on G (refer to eg. Boothby (1986) for details). For any point $g \in G$, $E_{i,g} f$ is the directional derivative of the function $f \in \mathcal{C}^\infty(G)$ in the direction of E_i. The vector $\sum_i (E_{i,g} f) E_{i,g}$ is the gradient vector of f at g. Now, by the existence theorem for ordinary differential equations, there exists a flow $\xi(t) \in G$ which is generated by the gradient field

$$\frac{d\xi(t)}{dt} = \sum_i \left(E_{i,\xi(t)} f\right) E_{i,\xi(t)} = \sum_i \left(E_{i,\xi(t)} f\right) \xi(t) E_i,$$

and thus

$$\xi(t) = \xi(0) \exp\left(t \sum_i \left(E_{i,\xi(t)} f\right) E_i\right),$$

where $\exp(\cdot)$ is the matrix exponential. This can be extended to stochastic flows by adding noise terms (Kunita, 1984), setting

$$\xi(t) = \xi(0) + \int_0^t \left(\sum_{i=1}^d \left(E_{i,\xi(s)} f\right) E_{i,\xi(s)} \mathrm{ds} + \sum_{i=1}^d E_{i,\xi(s)} \circ \mathrm{dW}_i(s)\right), \tag{18}$$

where $W_1(t), \ldots, W_d(t)$ are independent, standard Wiener processes and $\circ$ denotes the Stratonovich integral.

Since GL(3) has a "flat" geometry this simplifies, and we can use the gradient computed above. The Metropolis-Hastings algorithm becomes as follows. For each step, set

$$S' = S_t \exp\left(\frac{h}{2} \nabla \ln \pi\left(S_t \,|\, r\right) + \sqrt{h} \sum_{i=1}^d w_i E_i\right), \tag{19}$$

where the w_i are independent standard normals and $E_1, \ldots E_d$ is a basis for the Lie algebra $\mathfrak{gl}(3)$. The acceptance probability becomes

$$\alpha\left(S_t, S'\right) = \min\left\{1, \frac{\pi(S' \,|\, r)\, q(w')}{\pi(S_t \,|\, r)\, q(w)} \left| \exp\left(\frac{h}{2} \nabla \ln \pi\left(S_t \,|\, r\right) + \sqrt{h} \sum_{i=1}^d w_i E_i\right)\right|\right\}, \tag{20}$$

where $q(\cdot)$ is the density of the standard normal distribution.

The theory for constructing stochastic differential equations (SDEs) on manifolds can be found in eg. Kunita (1984). For examples on random sampling on curved manifolds, see eg. Piccioni and Scarlatti (1994) and Srivastava et al. (1999).

5 Examples and further work

As an example we have estimated locations of trees in range images of forest. This is motivated from target recognition applications, in which clutter objects such as trees poses a significant challenge. The most tractable way of dealing with clutter is to use Markov random field models (Zhu and Mumford, 1997) or pixel models based on projections of primitive geometric objects (Lee and Mumford, 1999), but computational aspects aside, the most natural way is to use physical models of clutter objects in the same manner as for targets. Furthermore, trees have highly variable shapes and might have very different sensor signatures depending on tree type and time of year. In that respect trees are suitable for the object model presented in this paper.

Fig. 2 shows estimated shape and location of a tree in a simple image. We used the template in Fig. 1 initially located in the centre of the imaged region. The parameters of the likelihood model were considered fixed, but the results did not vary much over a suitable range of parameters. The parameters of the Poisson object process were fixed at $\lambda_0 = 10^{-3}$ and $\lambda_1 = 5 \cdot 10^{-3}$.

To investigate the methods sensitivity to occlusion we considered estimation of the shape and location of a partially occluded tree. To simplify the computations we used a sequential algorithm, first locating the foremost tree, then the occluded one. This method requires good initial guesses of the positions. We used a crude, but effective method, assuming trees to have a fixed depth and then convolving the image with a three-dimensional Gaussian kernel. Fig. 3 shows estimated locations of the two trees. The results demonstrate that the method tackles occlusion, relying on prior information in the regions not observed.

The object recognition problem could be handled in full generality using the reversible jump MCMC as in Rue and Hurn (1999), assuming an unknown number of objects of varying type.

Acknowledgement

We thank Jinggang Huang and Ann Lee for providing the laser range data; and Ann Lee, Finn Lindgren and Håvard Rue for helpful comments.

The first author is grateful to the Division of Applied Mathematics, Brown University for its hospitality during his stay in the winter 1999/2000.

References

Besag, J. (1994). Discussion on the paper by Grenander and Miller, *Journal of the Royal Statistical Society, Series B* **56**: 591–592.

Boothby, W. (1986). *An Introduction to Differentiable Manifolds and Riemannian geometry*, Academic Press.

Christensen, O. F., Møller, J. and Waagepetersen, R. P. (2000). Geomteric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalised linear mixed models, *Technical report*, Department of Mathematical Sciences, Aalborg University, Denmark.

Gliklikh, Y. E. (1996). *Ordinary and stochastic differential geometry as a tool for mathematical physics*, Kluwer Academic Publishers.

Goodman, J. (1975). Statistical properties of laser speckle patterns, *in* J. Dainty (ed.), *Laser Speckle and Related Phenomena*, Springer Verlag, Berlin.

Green, T. J. and Shapiro, J. H. (1992). Maximum-likelihood laser radar range profiling with the expectation-maximization algorithm, *Optical Engineering* **31**(11): 2343–2354.

Green, T. J. and Shapiro, J. H. (1994). Detecting objects in three-dimensional laser range images, *Optical Engineering* **33**(3): 865–874.

Grenander, U. (1993). *General Pattern Theory*, Oxford University Press.

Grenander, U., Chow, Y. and Keenan, D. M. (1991). *Hands: a Pattern Theoretic Study of Biological Shapes*, Research Notes on Neural Computing, Springer, Berlin.

Hansen, M. B., Møller, J. and Tøgersen, F. A. (2000). Bayesian contour detection in a time series of ultrasound images through dynamic deformable template models, *Preprint*, Aalborg University, Denmark.

Husby, O. (2001). High-level models in ultrasound imaging, *Technical report*, Department of Mathematical Sciences, Norwegian University of Science and Technology.

Jacobs, S. P. and O'Sullivan, J. A. (1997). High resolution radar models for joint tracking and recognition, *Proc. IEEE National radar conference*, pp. 99–104.

Jacobs, S. P., O'Sullivan, J. A., Faisal, M. and Snyder, D. L. (1997). Automated target recognition system using high resolution radar, *Proc. Third workshop on conventional weapons ATR*.

Jain, A. K. and Dorai, C. (2000). 3D object recognition: Representation and matching, *Statistics and Computing* **10**(2): 167–182.

Koksal, A., Shapiro, J. and Wells, W. (1999). Model-based object recognition using laser radar range imagery, *Proc. SPIE*, Vol. 3718.

Kunita, H. (1984). Stochastic differential equations and stochastic flows of diffeomorphisms, *École d'Été de Probabilités de Saint-Flour*, Vol. XII -1982, Springer Vrlag.

Lanterman, A. D., Miller, M. I. and Snyder, D. L. (1997). General Metropolis-Hastings jump diffusions for automated target recognition in infrared scenes, *J. Opt. Engineering* **35**(4): 1123–1137.

Larsen, M. and Rudemo, M. (1998). Optimizing templates for finding trees in aerial photographs, *Pattern Recognition Letters* **19**(12): 1153–1162.

Lee, A. and Mumford, D. (1999). Scale-invariant random-collage model for natural images, *Proc. IEEE Workshop on Statistical and Computational Theories of Vision*, Fort Collins, CO.

Miller, M. I., Srivastava, A. and Grenander, U. (1995). Conditional-mean estimation via jump-diffusion processes in multiple target tracking/recognition, *IEEE Transactions on Signal Processing* **43**(11): 2678–2690.

Nair, D. and Aggarwal, J. K. (2000). Bayesian recognition of targets by parts in second generation forward looking infrared images, *Image and Vision Computing* **18**(10): 849–864.

Piccioni, M. and Scarlatti, S. (1994). An iterative Monte Carlo scheme for generating Lie group-valued random variables, *Advances in Applied Probability* **26**: 616–628.

Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin diffusions and their discrete approximations, *Bernoulli* **2**: 341–363.

Rue, H. and Hurn, M. A. (1999). Bayesian object identification, *Biometrika* **86**(3): 649–660.

Shapiro, J. H., Reinhold, R. W. and Park, D. (1986). Performance analyses for peak-detecting laser radars, *Proc. SPIE*, Vol. 663, pp. 38–56.

Srivastava, A. (1996). *Inferences on Transformation Groups Generating Patterns on Rigid Motions*, PhD thesis, Washington University, St. Louis, Mo.

Srivastava, A., Grenander, U., Jensen, G. R. and Miller, M. I. (1999). Jump-diffusion Markov processes on orthogonal groups for object recognition, *Preprint*, Department of Statistics, Florida State University, Tallahassee, FL.

Srivastava, A., Miller, M. I. and Grenander, U. (1997). Ergodic algorithms on special Euclidean groups for ATR, *Syst. and contr. in the 21th century*, Vol. 22, Birkhauser.

Webb, A. R. (2000). Gamma mixture models for target recognition, *Pattern Recognition* **33**(12): 2045–2054.

Zhou, D., Liu, G. and Wang, J. (2000). Spatio-temporal target identification method of high-range resolution radar, *Pattern Recognition* **33**(1): 1–7.

Zhu, S. C. and Mumford, D. (1997). Prior learning and gibbs reaction-diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(11): 1236–1250.

A Figures

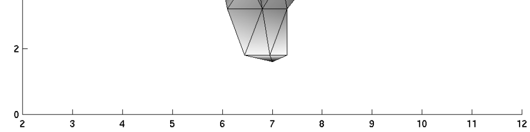
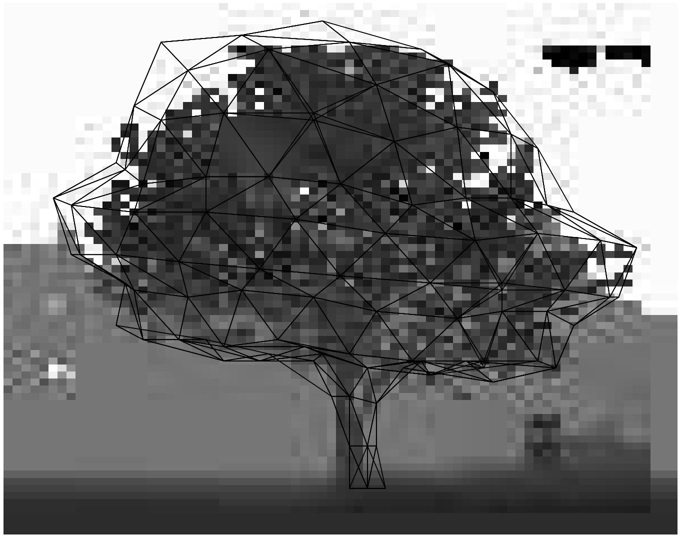
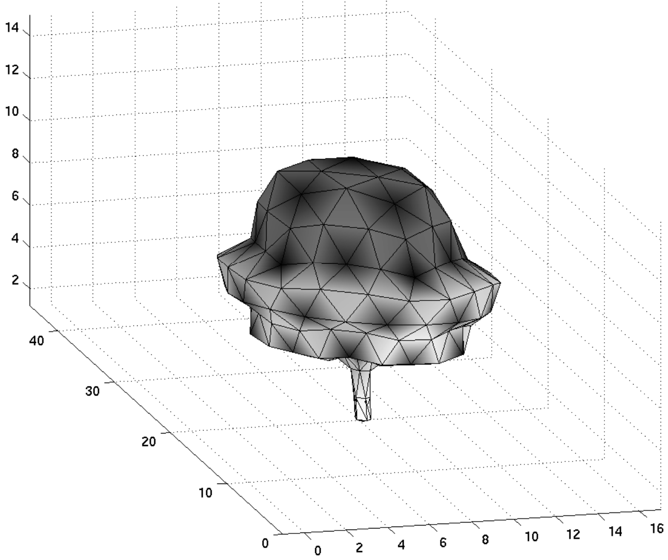


Figure 1: A tree template represented by a triangulated surface

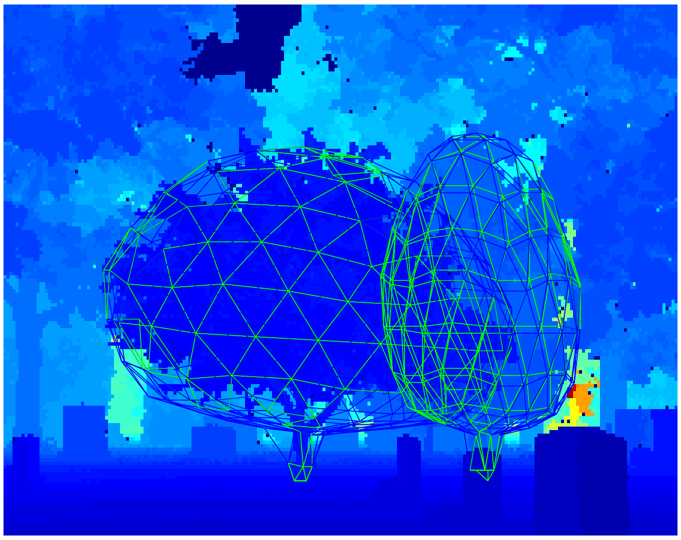


(a)

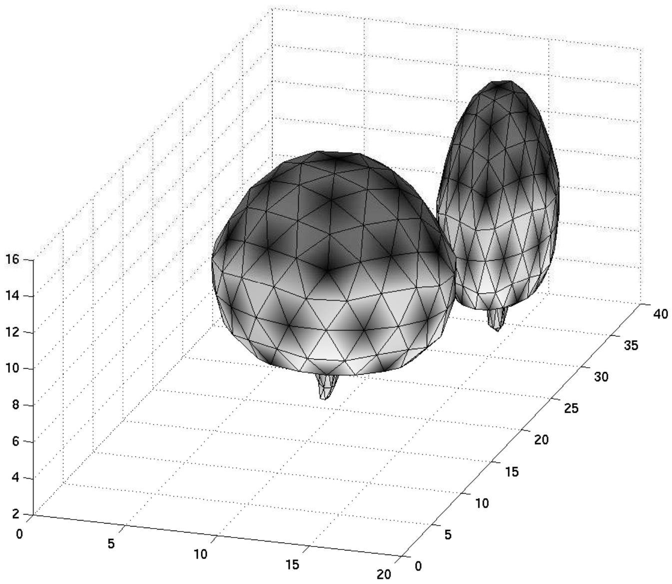


(b)

Figure 2: Estimated tree shape using the template in Fig. 1. Figure (b) shows the estimated location in space.



(a)



(b)

Figure 3: Estimation of occluded tree shape using the template in Fig 1. Figure (b) shows the estimated locations in space.