TOWARDS A THEORY OF NATURAL SCENES [1]

Ulf Grenander

January 2003

1 Problem Formulation.

Is it possible to build a science of natural scenes? A science that would make it possible to analyze scenes in a systematic manner, enabling us to classify,understand and predict their behavior. What we have in mind are methods that are sufficiently precise so that they can be implemented successfully by computer code. At first glance it appears unlikely that this would be possible; the variation in natural scenes is immense and it is difficult to see what sort of laws could govern their appearence. To borrow from history of science, it is hard to imagine laws like those of rational mechanics where deterministic, laws, differential equations, exist that enable us to understand and predict the way material bodies move, how they interact with each other. On the other hand, statistical mechanics offers a better paradigm, sacrificing an exact, detailed description of the systems in favor of a statistical representation that only makes probabilistic statement about their behavior. This is what David Mumford has suggested repeatedly, to measure the visual world as it appears to us macroscopically using statistical descriptions.

But how can we obtain such probabilities? Does it make sense to describe natural scenes in statistical terms, do they have enough stabilities and invariances for a theory to be feasible?

There is a growing literature dealing with this or related questions, mainly of empirical nature,and they have shown some remarkable regularities, see e.g. Huang, Mumford (1999). We shall try to derive analytical, model-based, results.

2 Marginal Probabilities for Natural Scenes.

Let us first recall some recent analytical results. The Transported Generator Model, TGM, introduced in Grenander, Miller, Tyagi (1999), defines configurations as

$$TGM : c = \{a_i s_i g^{\alpha_i}; i = ... - 1, 0, 1, ...\} \tag{1}$$

[1]This work has been supported by NSF DMS-00774276

with the i.i.d. amplitudes $a_i = N(0, \sigma^2)$, generator shapes $g^\alpha \in G^\alpha$ from an α index class of generator spaces G^α transported by a similarity group S. The random values of the similarity group elements are given by a homogeneous Poisson process (with respect to Haar measure) on the group. If the shapes are defined as real valued functions in the plane, a single α-class, and S is the translation group in the plane (1) reduces to

$$TGM : c = \{a_i g(x - x^i); i = ... - 1, 0, 1, ...; x = (x_1, x_2), x^i = (x_1^i, x_2^2) \in \mathbf{R}^2\} \tag{2}$$

With the identification rule "add" we get images

$$TGM : I(x) = \sum_{i=-\infty}^{\infty} a_i g(x - x^i); i = ... - 1, 0, 1, ...; x = (x_1, x_2), x^i = (x_1^i, x_2^i) \in \mathbf{R}^2 \tag{3}$$

REMARK 1: The TGM is related to the "dead leaves" and "random collage" models, Chi (1998),Ruderman (1997)

REMARK 2: The assumption about Gaussian distributions in the TGM can be relaxed as well as identifying images by the "add" rule. The latter can be replaced by the "min" rule leading to

$$TGM : I(x) = min_i[a_i g(x - x^i)]; i = ... - 1, 0, 1, ...; x = (x_1, x_i), x^i = (x_1^i, x_2^i) \in \mathbf{R}^2 \tag{4}$$

slightly more realistic. For a discussion of the pattern theoretic concepts used here see Grenander (1993).

We can make analytical statements about the images generated by the TGM (Grenander,Miller, Tyagi (1999):

THEOREM. *The 1D marginal distribution of $I(x)$ is infinitely divisible with non-negative kurtosis .*

This implies that the distributions are not Gaussian, which agrees with the empirical fact , often reported in the literature, that image ensembles of natural scenes usually appear non-Gaussian, see e.g. Huang,Mumford (1999). The histograms typically have a cusp at zero. This agrees with the analytical statement, Grenander (199a,b):

THEOREM. *Linear functionals of $I(\cdot)$ that annihilate constants have approximately marginal densities of the form*

$$f(x; p, c) = \frac{1}{Z(p, c)} x^{p-1/2} K_{p-1/2}(\sqrt{2/c}|x|) \tag{5}$$

where K is the modified Bessel function and the normalizing constant

$$Z(p, c) = \sqrt{(}\pi)\Gamma(p)(2c)^{p/2+1/4} \tag{6}$$

The Bessel K distributions are symmetric and unimodal for the mode at zero. For $p < 1$ they have a cusp at zero. For p = 1, f(x; p; c) is the density of a double exponential. As $p \to \infty$ they tend to Gaussian limits. The tails are heavier then Gaussian ones. In order to apply this to data we use the moments of the distributions of the given functional L.

THEOREM. *Using the relations*

$$E[(LI)^4] - 3\{E[(LI)^2]\}^2 = 3pc^2; E[(LI^2] = pc \tag{7}$$

and replacing the theoretical cumulants by their empirical analogs,i. e. method of moments estimation of the Bessel K parameters, we get the estimates

$$p^* = \frac{3}{sample\ kurtosis(I)}; c^* = \frac{sample\ variance(I)}{p^*} \tag{8}$$

Calculating these estimates for three images we get the theoretical Bessel densities (whole lines) displayed together with the histograms (dotted lines) in Figure 1. The amazing agreement of the Bessel K hypothesis with data that we see in this figure holds in most cases: *a* *universal law for natural scenes.*

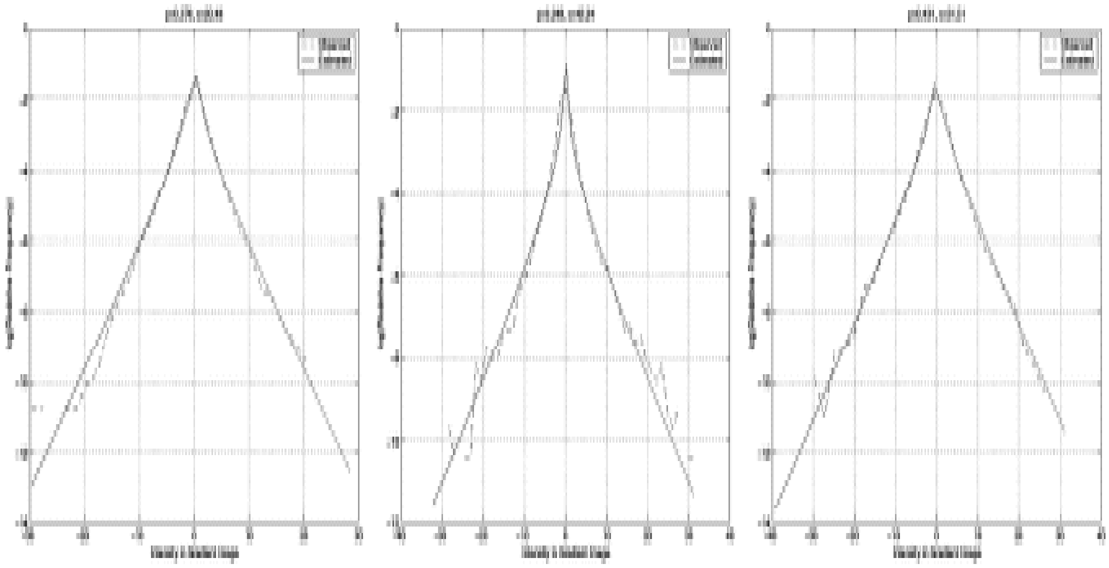

Figure 1

2.1 Higher Order marginals.

This success makes it natural to try to extend the result to higher dimensional marginals of filtered images, to find an approximation to the bivariate distribution of two stochastic variables $l_1 = L_1 I$ and $l_2 = L_2 I$, where L_1 and L_2 are two linear operators in $\mathcal{I}$ that annihilate constants. We shall complete the treatment as sketched in Grenander-Srivastave

(2001) using the Cramer-Wold device. Consider the characteristic function of the random variable $l = z_1 l_1 + z_2 l_2$

$$\psi(t) = E[exp(itl)] = E[exp(it < z, L >]; z = (z_1, z_2); L = (l_1, l_2) \tag{9}$$

But we know that the characteristic function is approximately

$$\psi_{approx}(t) = \frac{1}{[1 + t^2 c(z_1, c_2)]^{p(z_1,z_2)}} \tag{10}$$

Putting $t = 1$ this gives us approximately the bivariate characterristic function of the random vector L

$$\phi_{approx}(z) = \frac{1}{[1 + c(z_1, c_2)]^{p(z_1,z_2)}} \tag{11}$$

Now we use the relations for the Bessel K approximation

$$c = \frac{k_2}{3k_1}; p = \frac{3k_1^2}{k_2} \tag{12}$$

with the cumulants

$$k_1 = E[l^2]; k_2 = E[l^4] - 3\{E[l^2]\}^2 \tag{13}$$

This gives us

$$k_1 = z_1^2 E[l_1^2] + 2z_1 z_2 E[l_l l_2] + z_2^2 E[l_2 l_2] \tag{14}$$

$$k_2 = Z_1^4 E[l_1^4] + 4z_1^3 z_2 E[l_1^3 l_2] + 6z_1^2 z_2^2 E[l_1^2 l_2^2] + 4z_1 z_2^3 E[l_1 l_2^3] + z_2^4 E[l_2^4] \tag{15}$$

Using polar coordinates $z_1 = rcos\theta; z_2 = rsin\theta$ we introduce the trigonometric polynomials

$$P_1(\theta) = cos^2\theta E[l_1^2] + 2cos\theta sin\theta E[l_l l_2] + sin^2\theta E[l_2 l_2] \tag{16}$$

$$P_2(\theta) = cos^4\theta E[l_1^4] + 4cos^3\theta sin\theta E[l_1^3 l_2] + 6cos^2\theta sin^2\theta E[l_1^2 l_2^2] + 4cos\theta sin^3\theta E[l_1 l_2^3] + sin^4\theta E[l_2^4] \tag{17}$$

so that we can write

$$k_1 = r^2 P_1(\theta); k_2 = r^4 P_2(\theta) \tag{18}$$

Substituting this in equation (?) we get

$$\phi_{approx}(z) = \frac{1}{[1 + r^2 \frac{P_4(\theta}{P_2(\theta)}]^{\frac{P_2^2(\theta)}{P_4(\theta)}}} \equiv \frac{1}{[1 + r^2 a(\theta)]^{b(\theta)}} \tag{19}$$

To apply this to data we use the straigtforward estimates

$$E[l_1^2] \approx 1/n \sum_x (L_1 I)^2(x); E[l_l l_2] \approx 1/n \sum_x (L_1 I)(x)(L_2)(x); E[l_2^2] \approx 1/n \sum_x (L_2 I)^2(x) \quad (20)$$

and

$$E[l_1^4] \approx 1/n \sum_x (L_1 I)^4(x); E[l_1^3 l_2] \approx 1/n \sum_x (L_1 I)^3(x)(L_2 I); E[l_1^2 l_2^2] \approx 1/n \sum_x (L_1 I)^2(x)(L_2 I)^2(x)(21)$$

$$E[l_1 l_2^3] \approx 1/n \sum_x (L_1 I)(x)(L_2 I)^3(x); E[l_2^4] \approx 1/n \sum_x (L_2 I)^4(x) \quad (22)$$

We can apply (?) to range images by performing the 2D Fourier transform, the FFT, and get For $L1$ and $L2$ both being the discretized $\frac{\partial I}{\partial x_1}$, the latter translated horizontally

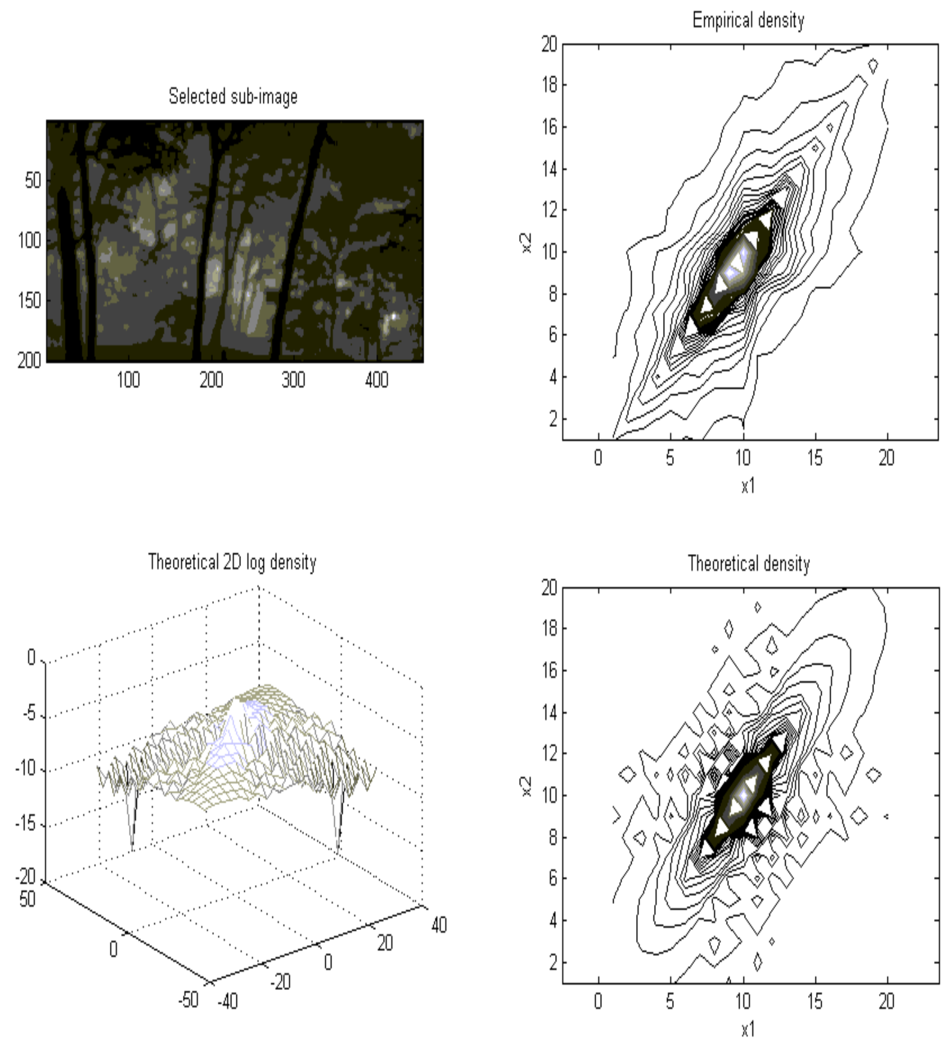

Figure 1a
The agreement is not as amazing as for the 1D marginals but still surprisingly good

considering the nature of the approximation. With $L1$ the same but $L2$ being $\frac{\partial I}{\partial x_2}$ we get

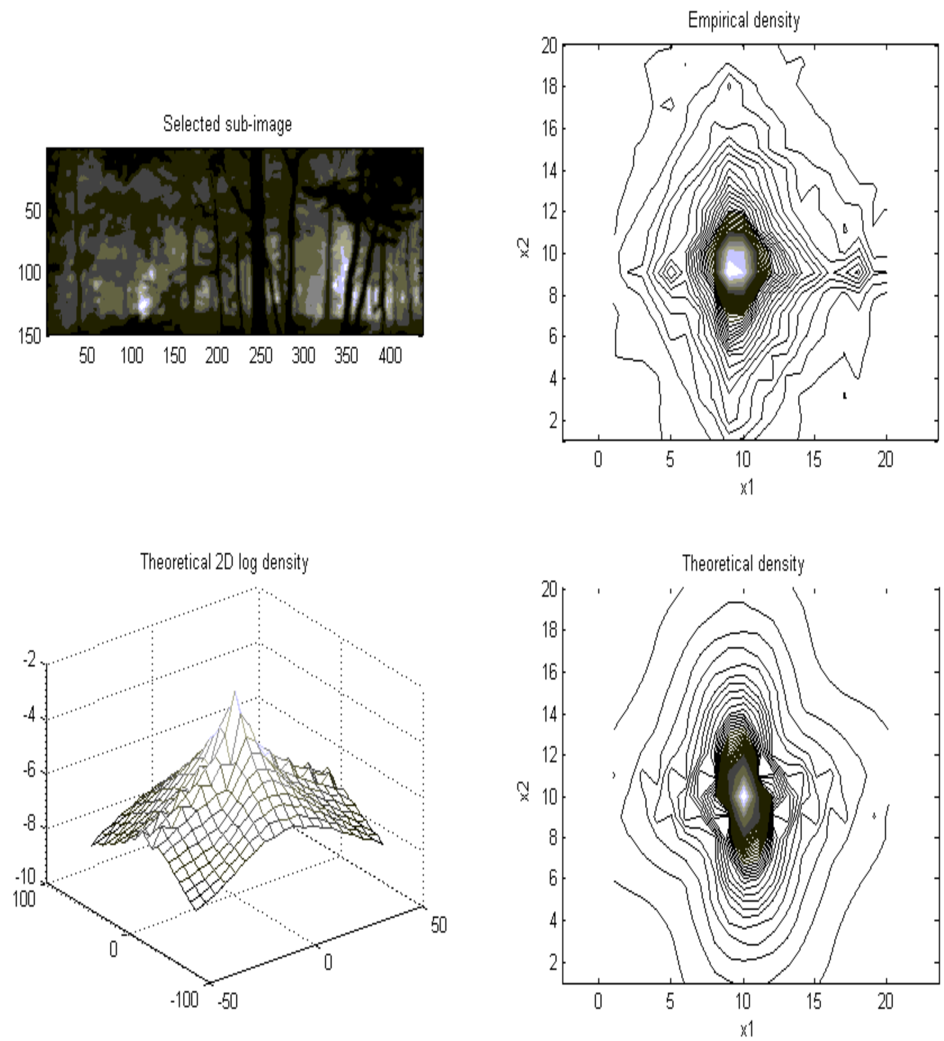

Figure 1b

and, not as good,

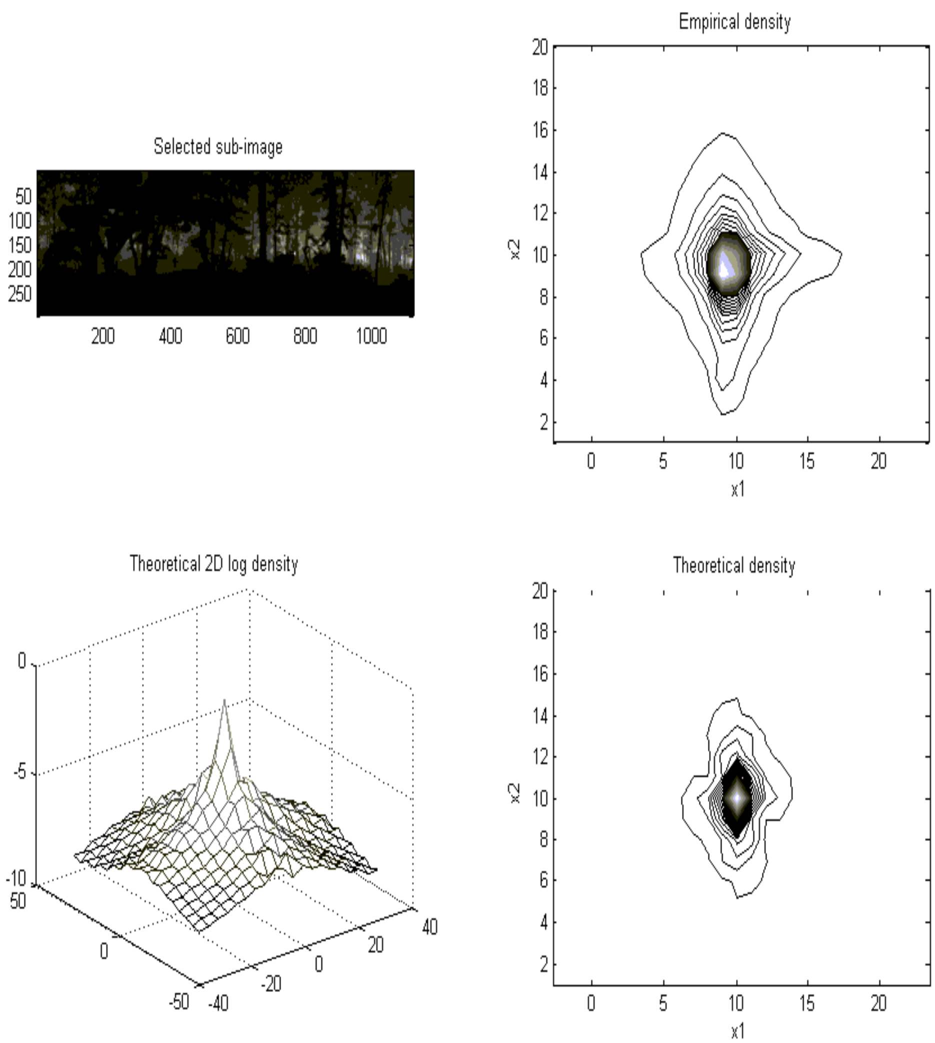

Figure 1c

This is better. However for a very heterogeneous image the approximation breaks down

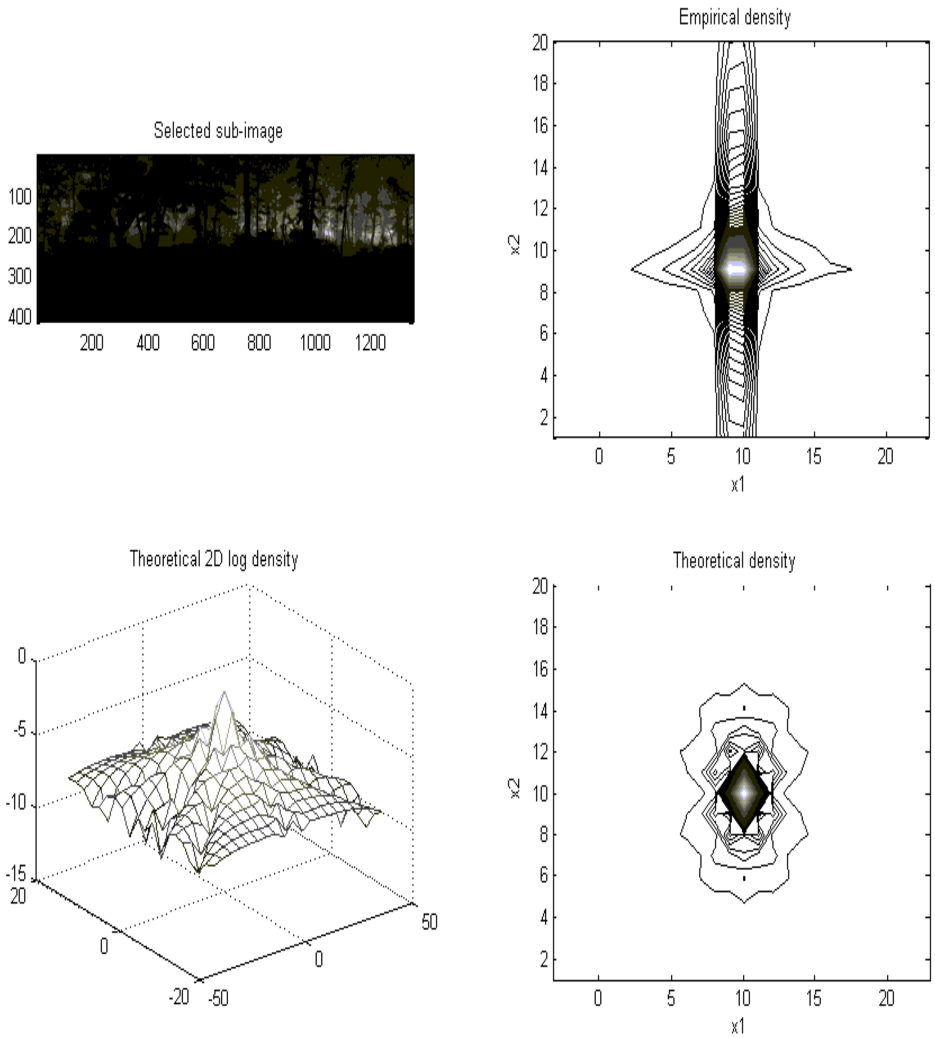

Figure 1d

2.2 Coarse Structure of 2D densities.

The trigonometric polynomials (P_2 and P_4 have a geometric interpretation. But first a short excursion in elementary probability theory. Say that a random 2-vector has the bivariate density $f(x) = f(x_1, x_2); x = (x_1, x_2)$ and the characteristic function written in polar coordinates $(x_1, x_2) \leftrightarrow (\rho, \alpha), (z_1, z_2) \leftrightarrow (r, v)$

$$g(z) = g(z_1, z_2) = \int_{x \in \mathbf{R}^2} exp[i < z, x >] f(x) dx = \int_{-\pi<\alpha<\pi} \int_{\rho=0}^{\infty} cos[\rho r cos(v-\alpha)] f(\rho, \alpha) \rho d\rho d\alpha \quad (23)$$

With some abuse of notation we let $f(\rho, \alpha)$ mean the same thing as $f(x_1, x_2)$, $g(r, v)$ the same as $g(z_1, z_2)$ Differentiate w.r.t. r twice and then put $r = 0$

$$(\frac{\partial^2 g(r,v)}{\partial r^2})_{r=0} = -\int_{-\pi<\alpha<\pi} \int_{\rho=0}^{\infty} f(\rho, \alpha) cos^2(v-\alpha) \rho^3 d\rho d\alpha \quad (24)$$

Introduce

$$F(\alpha) = \int_{\rho=0}^{\infty} f(\rho, \alpha) \rho^3 d\rho \quad (25)$$

so that

$$(\frac{\partial^2 g(r,v)}{\partial r^2})_{r=0} = -\int_{-\pi<\alpha<\pi} F(\alpha) cos^2(v-\alpha) d\alpha = -\pi \int_{-\pi<\alpha<\pi} F(\alpha) w(v-\alpha) d\alpha \quad (26)$$

with the weight function

$$w(u) = 1/\pi cos^2(u); \int_{-\pi<u<\pi} w(u) du = 1 \quad (27)$$

so that the second derivative is equal to $-\pi W[F(\cdot + v)]$, the w average of the length function F translated v steps.

Applying this to the function ϕ_{approx}, we get for the length function

$$W[F(\cdot + v)] = 2/\pi b(v) a(v) = 2/\pi P_2(v) \quad (28)$$

Hence maxima and minima of $F(\cdot)$ will approximately correspond to those of the trigonometric polynomial $P_2(\cdot)$. We illustrate this in Figure 1f where we see in the lower left panel the oblong shape oriented more or less as the empirical density in the upper right; this agrees with the above theoretical treatment.

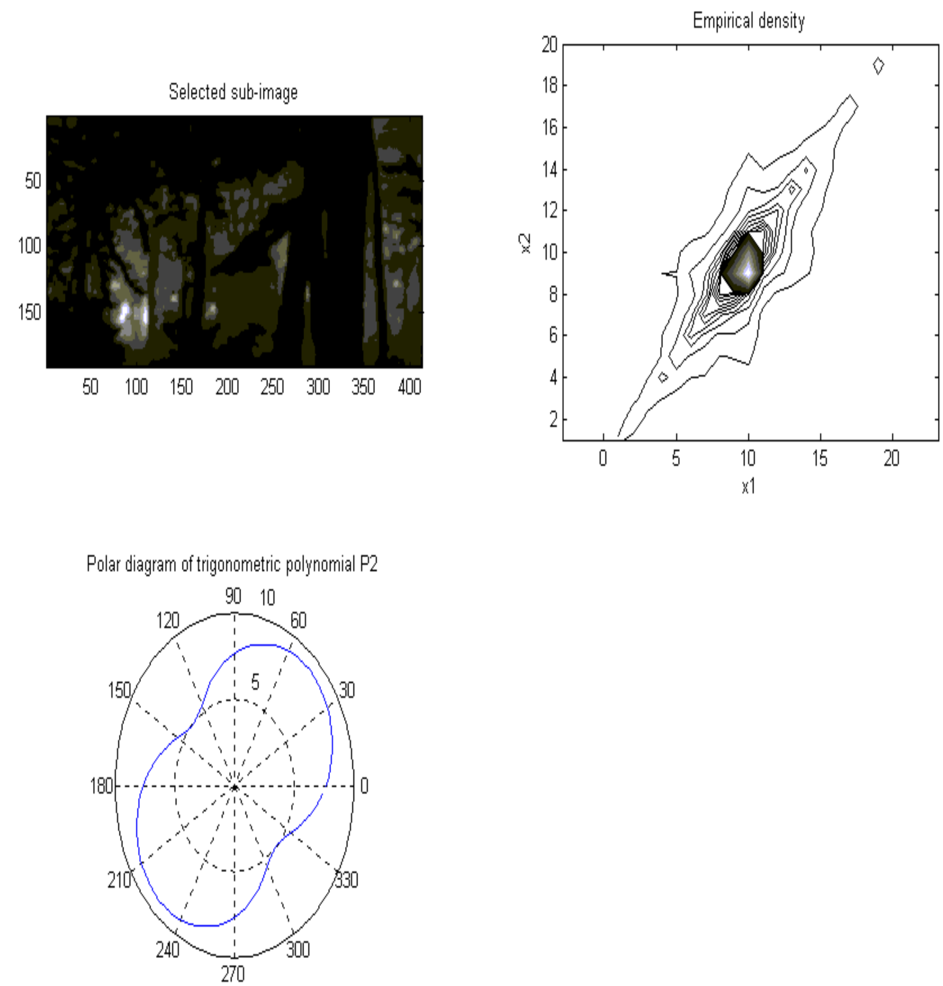

Figure 1e

In the next diagram with a large horizontal shift we are close to independence between L_1I and L_2I, correlation coefficient .04, and the oval shape shows no particular direction, it is almost a circle

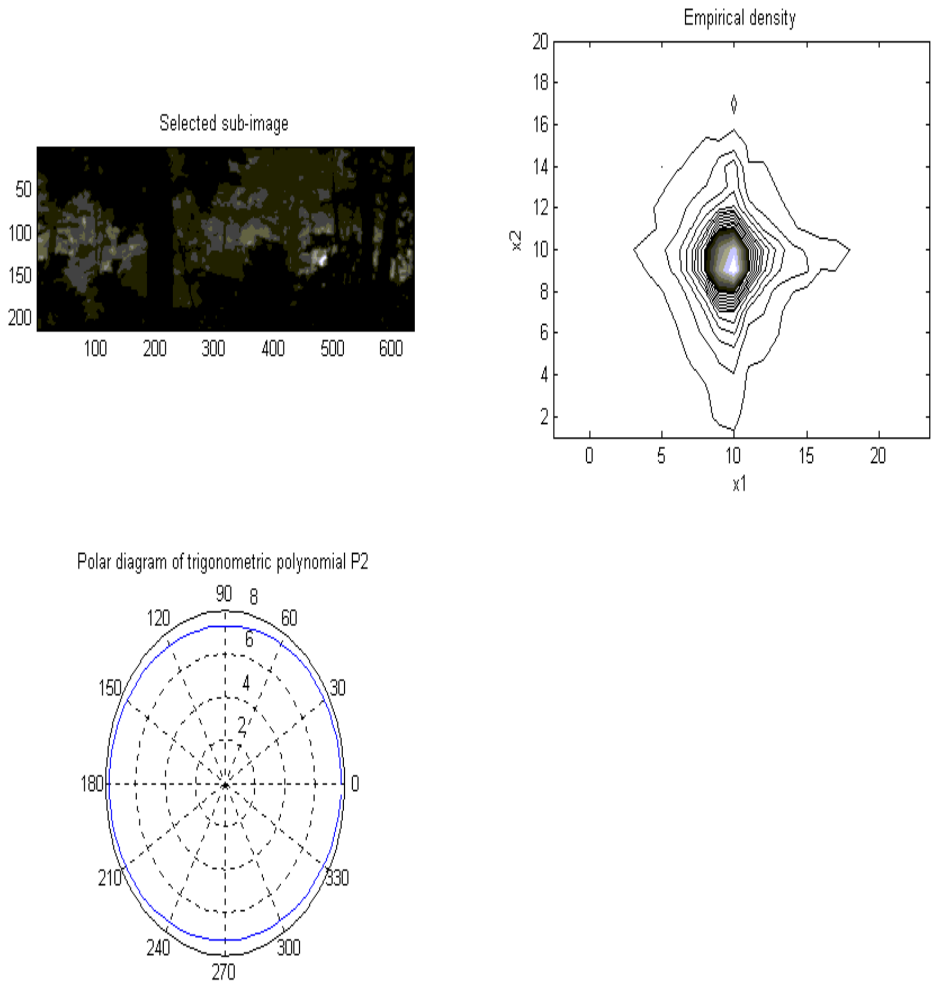

Figure 1f

Figure 1g has the main direction vertical

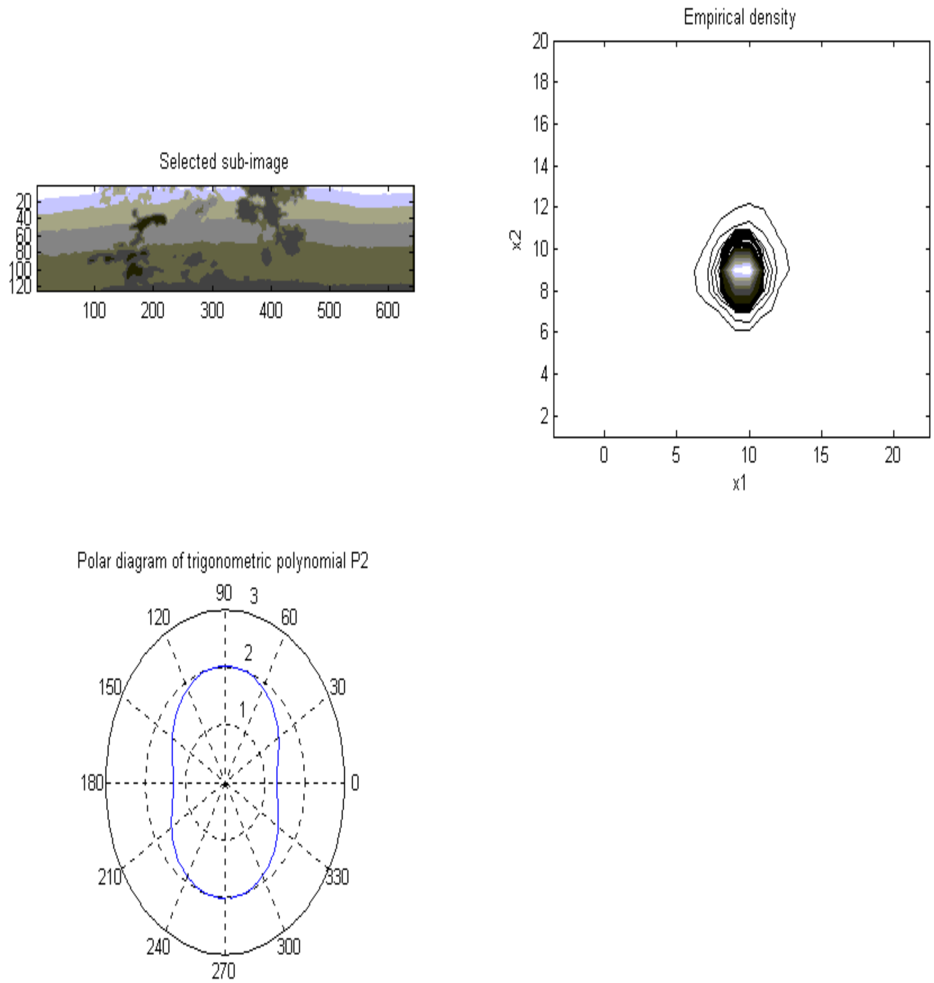

Figure 1g

It may be possible to extend this result to higher derivatives of G, perhaps getting more powerful results, but this has not been tried.

The success in 1D is remarkable considering the simple minded construction in the TGM, as is the limited success in 2D. The shapes are placed according a Poisson process in the plane and then observed with the addition identification rule (or the minimum rule). The significant feature of the model is that it considers the scene as made up of objects. The shape of the objects does not seem to matter except that they should have clearly delimited boundaries (Grenander (1999 a,b). But for higher dimensional marginals more knowledge *a priori* about the shapes is needed. The lack of this in the TGM is probably the reason why it does not work well for the 2D marginals of the functionals.

2.3 A full 3D model.

There is a need for firmer support for deriving of algorithms for the recognition of Objects Of Interest (OOI) hidden against a background of natural scene clutter, perhaps hiding part of the OOI. This is offered by the B3M

$$scene = \cup_\nu s_\nu g_\nu \tag{29}$$

with the objects represented by generator templates $g_\nu \in G^{\alpha_\nu}$; see GPT p.3, and, again, the s's form a stochastic point process over the group S. Here α is a generator index, see GPT p. 19, that divides the generator space into index classes

$$G = \cup_\alpha G^\alpha \tag{30}$$

The generators here mean the surface of the respective objects, and the index classes could represent different types of objects, trees, buildings, vehicles...We shall call the largest distance recorded by the range camera $range_{max}$.

In the case of range pictures it is natural to introduce a 3D polar coordinate system (r, ϕ, ψ) where r means the distance from the camera to a point in space and ϕ is the azimuth angle and ψ the elevation angle so that we have the usual relation

$$x_1 = rcos\phi cos\psi; x_2 = rsin\phi cons\psi; x_3 = rsin\psi \tag{31}$$

A point $x = (x_1, x_2, x_3)$ is transformed into Cartesian coordinates $u = (u_1, u_2)$ in the focal plane U of the camera by a perspective transformation that we shall call T. Hence the range image has the pixel values, in the absence of noise,

$$I(u) = min_\nu\{(Ts_\nu g^{\alpha_\nu})(u)\} \tag{32}$$

This version of the B3M will be used in Section 7. Denote by $support[Tsg)(\cdot)]$ the projected set of the OOI in the image plane, in other words where $(Tsg)(u) < range_{max}$.

It will be assumed that the $(Tsg)(u)$ is a C_2 function of (s, u) except of course at the projected boundary of the OOI, $\partial support[Tsg)(\cdot)]$. This regularity assumption will be used in Section 7, but may be extended to allow for a larger set of discontinuities of Lebesge measure zero.

We shall try to deepen our understanding of the structure of natural scenes by exploiting more knowledge about the shapes, the generators, that make up the scenes. But first we shall make precise what *a priori* information about the scenes is available to the observer, and what means of acquiring the images are being used. We emphasize this for the same reason that one emphasizes context in linguistic discourse.

3 Knowledge Status.

It would be a serious mistake to think of scene understanding as a problem with the observer equipped with objective and static knowledge about the world from which the scene is selected. On the contrary, the knowledge, codified into a prior measure, evolves over time and may be different for different observers. The well known visual illusions speak to this; the ambiguities are resolved completely only when additional information about the scene is made available.

Think of a person looking for an OOI in a landscape never before see by him - he will be capable of less powerful inference than some one familiar with the landscape. If we send out a robot to search for vehicles in a forest it is clear that it will perform better if equipped with an adequate map than it would otherwise. This is obvious, but with further reaching implications than may be thought at first glance.

The Hungarian probabilist Alfred Renyi used to emphasize that all probabilities are conditional. We believe in this, and conclude that any realistic approach to Bayesian scene understanding must be based on prior probabilities that mirror the current *information status*. The automatic search in a desert landscape for a vehicle using a photograph taken a day ago will be based on a prior different from the situation with no such photograph, just the knowledge that it is a desert. In the latter case the prior may be a 2D Gaussian stochastic process with parameters characteristic for deserts in that part of the world. In the first the prior may be given via a map computed from the photograph superimposed with another Gaussian processing representing possible changes in the location of the sand dunes during the last day; obviously a situation more favorable for the inference machine.

Other incidentals that could/should influence the choice of prior are, meteorological conditions, observed or predicted, position of sun, type of vegetation, topographical features known or given statistically, presence of artifacts like camouflage, ... For each likely information status we should build knowledge informations of the resulting scenes. This is a tall order, a task that will require mathematical skills and subject matter insight. It should be attempted and it will be attempted!

Any adequate formulation of the problem requires a careful description of the prior knowledge about the scenes and of the means available for observing it: the information status. We have argued elsewhere that this deserves our attention; here we shall elaborate on this view and organize the information in explicit terms. Since knowledge is structured information we shall speak instead of the *knowledge status* of the problem. The following discussion could be classified as *mathematical epistemology.*

We shall formalize the knowledge status as follows:

The knowledge status will be repreesented by a Knowledge Box:

$$\mathbf{K} = \{k_1, k_2, k_3, ...\} \tag{33}$$

with the knowledge elements $k_1, k_2, k_3, ...$

The k_i's can be deterministic or probabilistic descriptions of knowledge available to the algorithm for understanding. The set of K's used in a situation will be called the $\mathcal{K}$-lattice; it allows the lattice operations $K' \vee K''$ (increase of knowledge, includes sensor fusion) , and $K' \wedge K''$ (decrease of knowledge, loss of sensor input) and then the sup and inf operations. A partial order is naturally induced on $\mathcal{K}$; its operation will be denoted by the symbol $<$.

3.1 Examples.

Consider the following *Knowledge Box*

$\mathbf{K^{(1)}} =$

knowledge element	element descriptor
k_1	output from specified range camera
k_2	scene type forest; parameters=a,b,c...
k_3	tree type deciduous; parameters $\alpha, \beta, ...$
k_4	slowly rolling landscape;parameters=k,l...

With another technology we get another Knowledge Box

$\mathbf{K^{(2)}} =$

knowledge element	element descriptor
k_1	output from specified FLIR camera
k_2	intelligence: a vehicle is likely in the scene, parameters d,e,...

Combining both, $\mathbf{K} = \mathbf{K^{(1)}} \vee \mathbf{K^{(2)}}$, we get the knowledge status that will be assumed in the section 6. Still another:

$\mathbf{K^{(3)}} =$

knowledge element	element descriptor
k_1	intelligence: a vehicle in the scene can possibly be a tank
k_2	tank specification through a CAD representation

and the union $K^{(4)} = K \wedge K^{(3)}$.

4 World Model= Generators+Connectors+Priors .

The organization of algorithms for understanding natural scenes will be based on a *theory of the world* being observed. Indeed, without any theory predicated it seems impossible to organize and analyze the received images in a meaningful way. We shall express the theory in pattern theoretic form, see GPT [2], PART I.

[2]Refers to Grenander (1993)

4.1 Generators.

These are the primitives in terms of which the understanding will be organized. They belong to the knowledge lattice of the situation.. They often appear on different levels of specificity as shown in the following examples, as ordered on the $\mathcal{K}$ lattice.

EXAMPLES :

$trunk < trunk\ sides < curved\ trunk\ surface$
$foliage\ profile\ < detailed\ foliage < detailed\ foliage\ with\ holes$
$horizontal\ ground < linear\ slope\ ground < curved\ ground$
sky
all with range information.

4.2 Invariance via Similarity Groups.

Patterns are formed as equivalence classes of images w.r.t. a similarity group S; $\mathcal{P} = \mathcal{I}/S$. The most obvious similarities are in the space/time domain, say

(a) SE(3)= the special Euclidean group in $\mathbf{R}^3$ for change in location and pose
(b) G(3)= the Galilean group in $\mathbf{R}^3 \times \mathbf{R}$ also with motion
(c) A(3) = the affine group in $\mathbf{R}^3$ for change in location, pose and includes skewing
(d) D(3) = the group of diffeomorphisms in space for topological transformations

4.3 Connectors.

They connect some generators together following the regularity rules. They can be probabilistic in nature.

EXAMPLES :
$trunk\ side\ 1 \leftrightarrow trunk\ side\ 2$
$vehicle\ body \leftrightarrow wheel$
$wheel \leftrightarrow ground$

4.4 Probabilities.

Any realistic inference theory for natural scenes, must be probabilistic to account for the high variability in the observed images.

To fix ideas let us discuss natural scenes of forest type. For generators such as trunks, it makes sense to understand probability distributions of diameters in the standard frequentist sense, and to estimate them by examining many forest scenes and, for example, to measure the diameters at different heights. Or to describe their location on the ground level by some point process in the way studied in depth in the pioneeering work of Matern (1960). The probability that a tree in a given forest is pine or oak can also be obtained from measurements specifying the knowledge status, e.g. by a statistical map.

5 Interpolation.

We shall now study interpolation inference for forest scenes and use the knowledge box in Section 5. Having observed a range image $I^{\mathcal{D}}_{total}$ we use the FLIR knowledge in $K^{(3)}$ to select a sub-image $I^{\mathcal{D}}$ of size $l1 \times l2$ a a candidate for a region that may contain an OOI (Object Of Interest). The rationale behind this is that the FLIR has fairly low resolution so that it gives the position of the OOI with little accuracy, in contrast to the laser radar. Hence the information from the FLIR only gives us a confidence region as rectangle, *an attention field AF*. Within this rectangle we know the position with some probability distribution, uniform or not. Put a frame of width $(L1, L2)$ around the sub-image so that we have a somewhat bigger image $I^{\mathcal{D}}_{ext}$ of size $L1 \times L2$; see Figure 2 We shall use the information in the frame $F = I^{\mathcal{D}}_{ext} \setminus I^{\mathcal{D}}$ to interpolate the inner image $I^{\mathcal{D}}$, treating it as unknown. The reason for this is that we do not know if it contains an OOI, and if it does, where is it and what is its pose ? To answer such a question we must know something about the background and that is exactly the task of the interpolator.

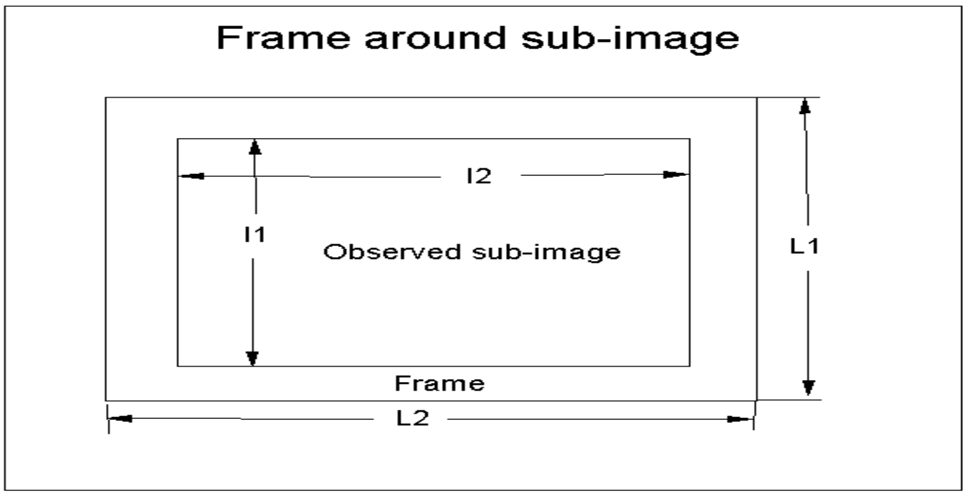

Figure 2

5.1 Non-specific method.

Let us attempt interpolation schemes based on minimizing the conditional energy derived from Bessel K distributions. Assume that the joint density of the images has the form

$$f(I) = \frac{1}{Z} \prod_{i_1=1,i_2=1}^{l_1,l_2} b([I^{\mathcal{D}}(i_1+1,i_2) - I^{\mathcal{D}}(i_1,i_2); c, p] b[I^{\mathcal{D}}(i_1,i_2+1) - I^{\mathcal{D}}(i_1,i_2); c, p] \qquad (34)$$

using the available boundary values. The implicit independence assumption in this formula is of course not valid, but is introduced in the spirit of mean field theory. Hence

$$E_{cond} = \sum_{i_1=1}^{l_1} \sum_{i_2=1}^{l_2} e[I^{\mathcal{D}}(i_1,i_2)] \qquad (35)$$

to be minimized over all $I^{\mathcal{D}}(\cdot,\cdot))$ with boundary values BV obtained from the framed values and with the Laplacian

$$e[I^{\mathcal{D}}(i_1,i_2)] = f[I^{\mathcal{D}}(i_1+1,i_2)] + f[I^{\mathcal{D}}(i_1-1,i_2)] + f[I^{\mathcal{D}}(i_1,i_2+1)] + f[I^{\mathcal{D}}(i_1,i_2-1)] - 4f[I^{\mathcal{D}}(i_1,i_2)] \qquad (36)$$

We used the Bessel K densities b and

$$f(x) = log[b(x, c, p)] \tag{37}$$

referring to the expression (5). Note that this is approach is non-specific in that it does not specify the detailed pattern theoretic structure of the images in terms of generators and so on.

For the Gaussian case $p \to \infty$ we get interpolation with the classical harmonic function and their boundary value problem. For finite p-values it should be noticed that the minimum is not unique for $p \leq 1$: the energy is not convex for $p < 1$ and not strictly convex for $p = 1$; hence uniqueness of the minimum is not guaranteed,.

The results are disappointing. For classical harmonic function interpolation, in which each iteration replaces a pixel value by the mean of its foiur neighbors, we get, as the best result,

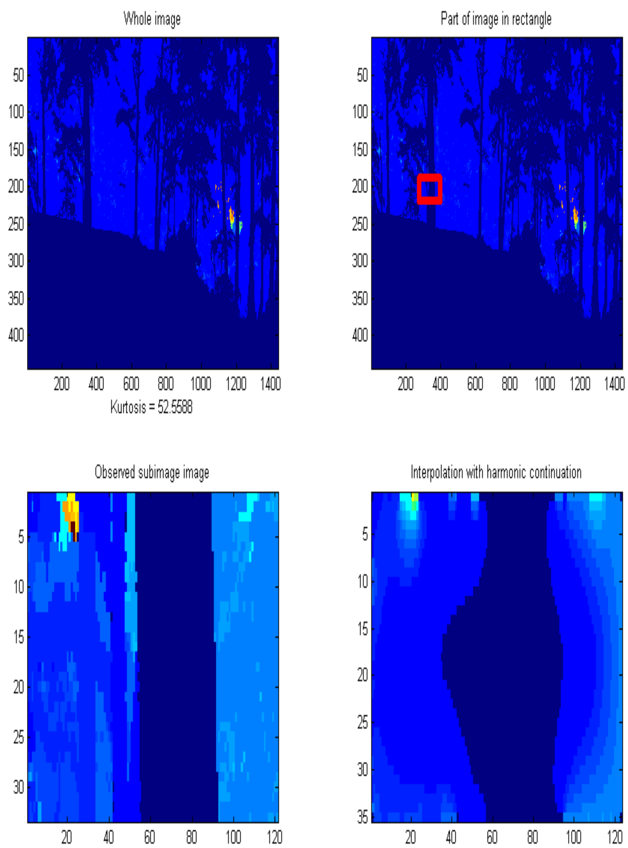

Figure 3

For $p = 1$, which means $l1$ norm and each iteration replaced a pixel value by the median of its four neighors, the interpolation performs even worse:

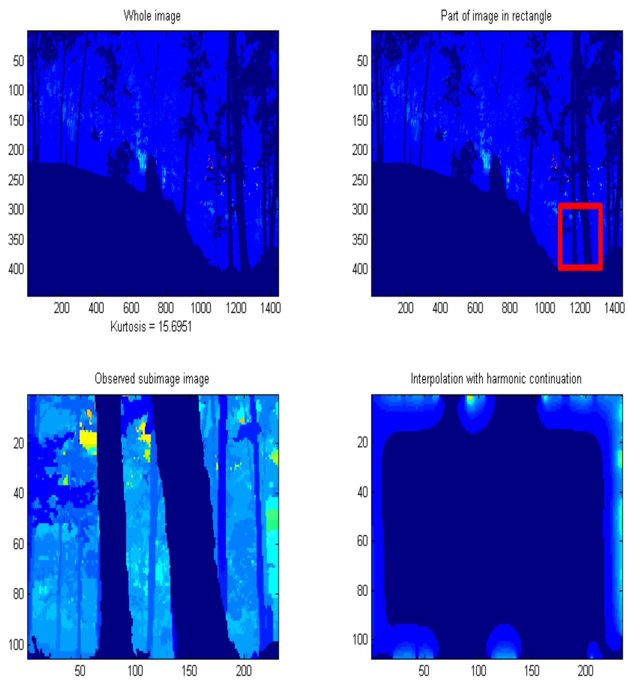

Figure 4

The reason for this poor inference performance is of course that the assumed prior probabilities do not catch much of the real image structure. Indeed, it says only that all the differences

$$I^{\mathcal{D}}(i_1+1, i_2) - I^{\mathcal{D}}(i_1, i_2), I^{\mathcal{D}}(i_1, i_2+1) - I^{\mathcal{D}}(i_1, i_2) \tag{38}$$

are i.i.d with Bessel K marginals conditioned by boundary values. This expresses the fact that the images are made up of objects - almost constant values over individual objects with jumps between them. But it does not say anything about the form of the objects. The knowledge status is too weak! Only 2D marginal distributions of the $I^{\mathcal{D}}$ are described . Of course we have also derived 3D approximations, but while the 2D Bessel K approximations provided highly accurate quantitative agreement with data, this was not the case for higher dimensions; only qualitative similarities with data were observed.

To get better results we must strengthen the knowledge status to specify the pattern structure of real forest pictures. But how much? Only as much as is necessary for reasonable inference performance, otherwise we can expect too slow algorithms and possibly overfitting the data. Although algorithmic speed is not our main concern at this stage, we shall aim at least for computational feasibility on current PC's.

5.2 Specific Method.

We shall use the generators from section 2.1: $G = \{foliage, ground, trunk, sky\}$. This generator space is a bare minimum and should be increased, but it will have to suffice for the present. To recognize these four type of generators we shall introduce indicators as follows.

First, compute the boundary value function $BV(s), s \in 1 \leq l$ along the boundary of $\partial I^{\mathcal{D}}$ with the arc length $l = 2l_1 + 2l_2 + 4$. A typical example is shown in Figure 5 Note how stretches of nearly constant range values or linearly increasing ones are separated by rapidly changing values.

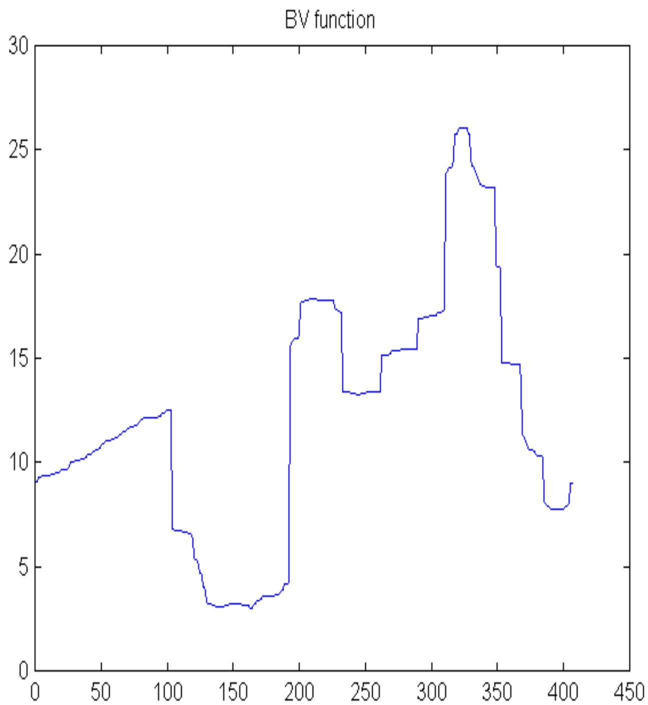

Figure 5

Now determine

$$M = max_s BV(s); m = min_s BV(s) \tag{39}$$

and introduce the range levels

$$[r_k, r_{k+1}]; k = 1, 2, ...N; r_k = m + (k-1)/N(M-m) \tag{40}$$

for some moderate natural number N and with the average range levels $m_k = (r_k + r_{k+1})/2$. This leads to intervals of the form $\{s : r_k \leq BV(s) < r_{k+1}\}$. Then reject small intervals with lengths less than some threshold value and filling in holes of length smaller than some other threshold. This gives us a set INTERVALS $= \{int_1, int_2, int_3, ...int_{n_{int}}\}$ of these intervals, each interval $int \in INTERVALS$ associated with some range level and will be denoted as $[p_1(\nu), p_2(\nu)]$ on the boundary of $I^{\mathcal{D}}$. We shall use modular addition $n_{int} + 1 \equiv 1$.

We shall need the following concept. Define a function

$$Q(s) = 1, s \in BELOW; Q(s) = 2, s \in RIGHT; Q(s) = 3, \in UP; Q(s) = 4, s \in LEFT \quad (41)$$

where $BELOW, RIGHT, UP, LEFT$ mean the four sides of the domain X of $I^{\mathcal{D}}$. Now introduce indicators. For each interval int_ν define the indicator

$$\omega_\nu^1 = LS(\nu) = line\ segment\ p_1(\nu) \rightarrow p_2(\nu) \quad (42)$$

in the rectangle X. The second class of indicators is more involved.

Consider a line segment $LS(\nu)$ with associated average range r_k and the corresponding s-set $S(\nu)$ along the boundary ∂X. In the frame F find all pixels $(i_1, i_2) \in F$ with range values in the interval int_ν. Among those pixels we find the ones connected with points in $S(\nu)$ according to the closest neighbor topology. In other words, find the topological component $C(\nu)$ that includes the set $S(\nu)$. Now find unit vectors U_1, U_2 to $p_1(\nu)$ and $p_2(\nu)$ respectively enveloping most of C as in Figure 6

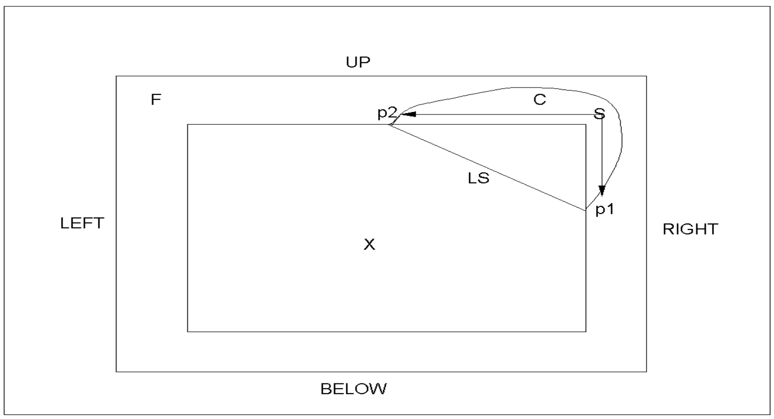

Figure 6

The precise definition of these vectors is given in the MATLAB software. We do not insist on that particular choice, others may be better and we leave this till later. Then we get indicators of the second class

$$\omega_\nu^2 = (U_1, U_2) \tag{43}$$

The rationale behind this choice of indicators is the following. To get a likely continuation of the line segment, the chord that cuts the foliage object, we shall use the directions indicated by the frame picture. In other words, we are estimating derivatives, gradients, which is a notoriously sensitive task for complex pictures with much local variability. Anyway, this is the purpose of the vectors U_1, U_2.

With the set of these indicators, $\Omega = \{\omega^1(\nu), \omega^2(\nu); \nu = 1, 2, \ldots\}$, we are ready to organize the inference. But first a detail caused by an artifact in the range data. The output of the particular laser radar used in Huang-Lee (199) is 0 for very large distances. To compensate for this we put

$$I^{\mathcal{D}}(i_1, i_2) = 0 \; redefined \; as \; I^{\mathcal{D}}(i_1, i_2) = M \tag{44}$$

but keep the unmodified $I^{\mathcal{D}}$ as $I^{\mathcal{D}}_{save}$.

5.3 Trunk.

For a line segment $LS(\nu)$ with $Q[p_1(\nu)] = Q[p_2(\nu)] = 3$, that is the segment belongs to UP, we calculate

$$r_{parallel} = R(\|U_1 - U_2\| \leq \epsilon_1) \tag{45}$$

with $R(x) = exp(-x)$ and

$$r_{vertical} = R(\|(U_1 + U_2)/2 - col(1,0)\| \leq epsilon_2) \tag{46}$$

We shall interpret the value r_1 as the heuristic probability that the two unit vectors are almost parallel. The value r_1 is the heuristic probability that the average vector $(U_1 + U_2)/2$ is almost vertical. We put

$$\phi_1(I^{\mathcal{D}}) \rightarrow \text{"}trunk\ element\ at\ (p_1(\nu) + p_2(\nu))/2\text{"} \tag{47}$$

With probability $r_1 r_2$ we introduce a partial interpolator image I_t^* with pixel values equal to M everywhere except at pixels belonging to a vertical rectangle extending from the interval int_ν downwards all the way to ∂X where they will be equal to the average range value r_k of $int(\nu)$. With the complementary probability we do nothing.

The form of R has been selected somewhat but not wholly arbitrarily. The index t is updated $t \rightarrow t+1$ for each decision to recognize a generator.

5.4 Foliage.

To infer foliage generators requires a new concept, *set continuation*. For an arbitrary interval $int(\nu) \in INTERVALS$ let us introduce another partial interpolator image I_t^*, now with all pixel values equal to M, except those in a set $CONT[SL(\nu)]$ where the pixel values shall be equal to the average range value r_k associated with $int(\nu)$. The set $CONT[SL(\nu)]$, the set continuation of the line segment $LS(\nu)$ with directions U_1, U_2, will be derived as follows.

To contine a set representing some object (here foliage of a tree), we should specify the knowledge status. It will depend upon what, if anything, is known about the tree species *a priori* - is it an oak or a pine... ? Say that we only know that its profile is "rounded" with piecewise continuous curvature. Real tree images have holes as modelled in Husby and Grenander (2001) but this fact will not be includeded in the current knowledge status. We shall think of the (discrete) set as made up of line segments with slowly varying positions and lengths, see Figure 7 and described below, forming a Markov process of order 2. The process continues as long a the line segments have positive length. We shall use the observables

p_1, p_2, U_1, U_2.

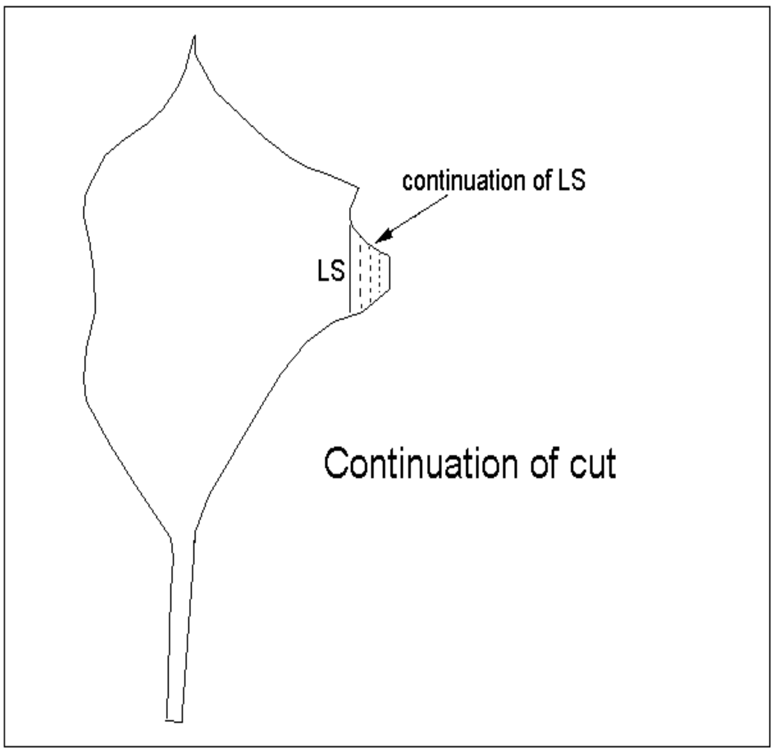

Figure 7

Say the cut is along $x_1 = constant$ with endpoints $[x_2^1(t), x_2^2(t)]$ on level $x_1 = t$. Let us assume, in terms of discrete space, that the cuts form a Markov process satisfying the coupled Langevin SDE's

$$x_2^1(t) = 2x_2^1(t-1) + x_2^1(t-2) - k * [x_2^1(t-1) - x_2^2(t-1)] + e_1(t) \tag{48}$$

$$x_2^2(t) = 2x_2^2(t-1) + x_2^2(t-2) + k * [x_2^1(t-1) - x_2^2(t-1)] + e_1(t) \tag{49}$$

with $e_1(t), e_2(t)$ meaning white noise $N(0, \sigma^2)$ and k is a restoring force coefficient. The mechanical analog of this means two mass points attracting each other with a force proportional to their distance and subject to random impacts. Initial condition $x_2^1(0) = p_1, x_2^2(0) = p_2, p_2 < p_1; x_2^1(1) = q_1, x_2^2(1) = q_2$, meaning that we specify boundary values of first order discretized. The equation should terminate as soon as $x_2^1(t) < x_2^2(t)$. In the mechanical analog $[q_1 - p_1, q_2 - p_2]$ is the initial velocity vector.

In Figure 8 we show some set continuations using this *Markov cut model*; primary cut in red, continuations in blue. Note how the direction tendencies at the initial cut propagate into the protuberance but gradually die out. The speed of this depends upon k that regulates the (random) size of the protuberance, while σ^2 controls the smoothness of the boundary.

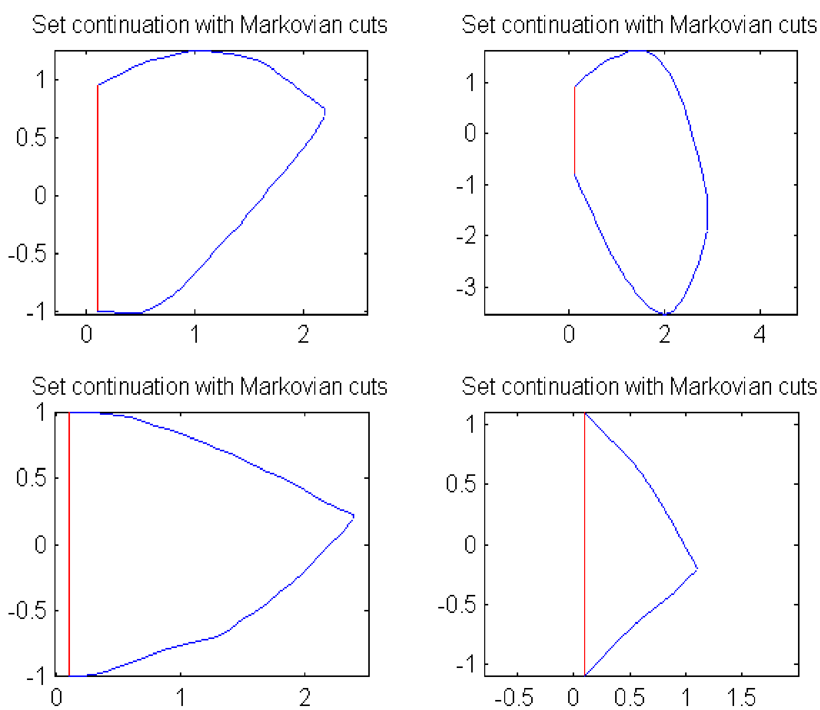

Figure 8

We then introduce a partial interpolator image I_t^* covering the continued foliage set. This concludes the discussion of the set continuation we shall use for foliage sets.

5.5 Ground.

Let us consider two intervals $i_1 = int(\nu_1), i_2 = int(\nu_2) \in INTERVALS$ with $Q(i_1) = 1 \; or \; 2, Q(i_2) = 4 \; or \; 1$ and associated with the same average range value V. This gives us four points $P_1 = p_1(\nu_1), P_2 = p_2(\nu_1), P_3 = p_1(\nu_2), P_4 = p_2(\nu_2)$. Find $U_1(\nu_1), U_2(\nu_1)$ for the interval i_1 and $U_1(\nu_2), U_2(\nu_2)$ for the interval i_2. Also the heuristic probability

$$r_{same\ height} = R(\|P_2 - P_3\| \leq epsilon_3 \; and \; \|P_4 - P_1\| \leq \epsilon_3) \tag{50}$$

where R is some heuristic probability measure. With probability $r_{same\ height}$ we then introduce a partial interpolator image I_t^* for the ground generator; I_t^* consisting of the quadilateral $P_2 \to P_3 \to P_4 \to P_1$ associated with the range V.

5.6 Sky.

For the intervals in INTERVALS find the ones associated with the range value 0, large distance. Apply the above set continuation method to extend the lines segmens represented by these intervals. In the union of these sets make the range equal to $range_{max}$, the largest distance that the laser radar outputs as a positive number.

Now we combine all the partial interpolator images to the interpolator

$$I^* = min_{t \in T} I_t^* \tag{51}$$

where T is the set of t-values that have been introduced successively in the procedure described in this section.

Let us look at some of the results.

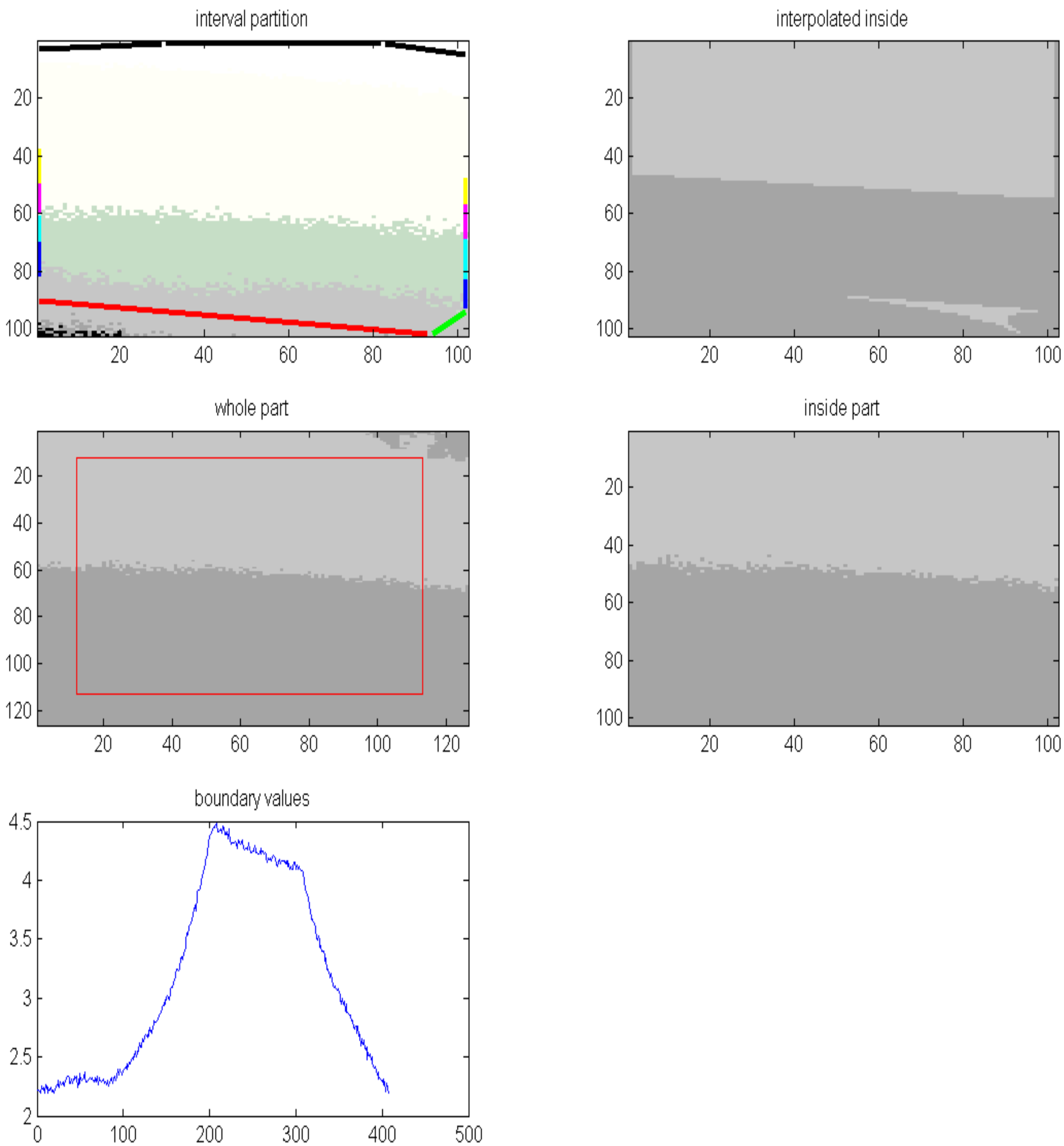

Figure 9

Figure 9 shows a simple task where the scene is simply a hill landscape with no other features. The upper left panel shows the selected sub-image and the middle left also the frame; the first one indicates how the algorithm has identidfied and organized boundary intervals into reasonable intervals. The boundary values are in the lower left panel as a function of arc length along ∂X. The upper right panels is the inference in the form of an interpolation while the middle right one shows the true image to be interpolated.

Obviously the algorithm does a fine job here, but the task was not really difficult enough. Now look at Figure 10

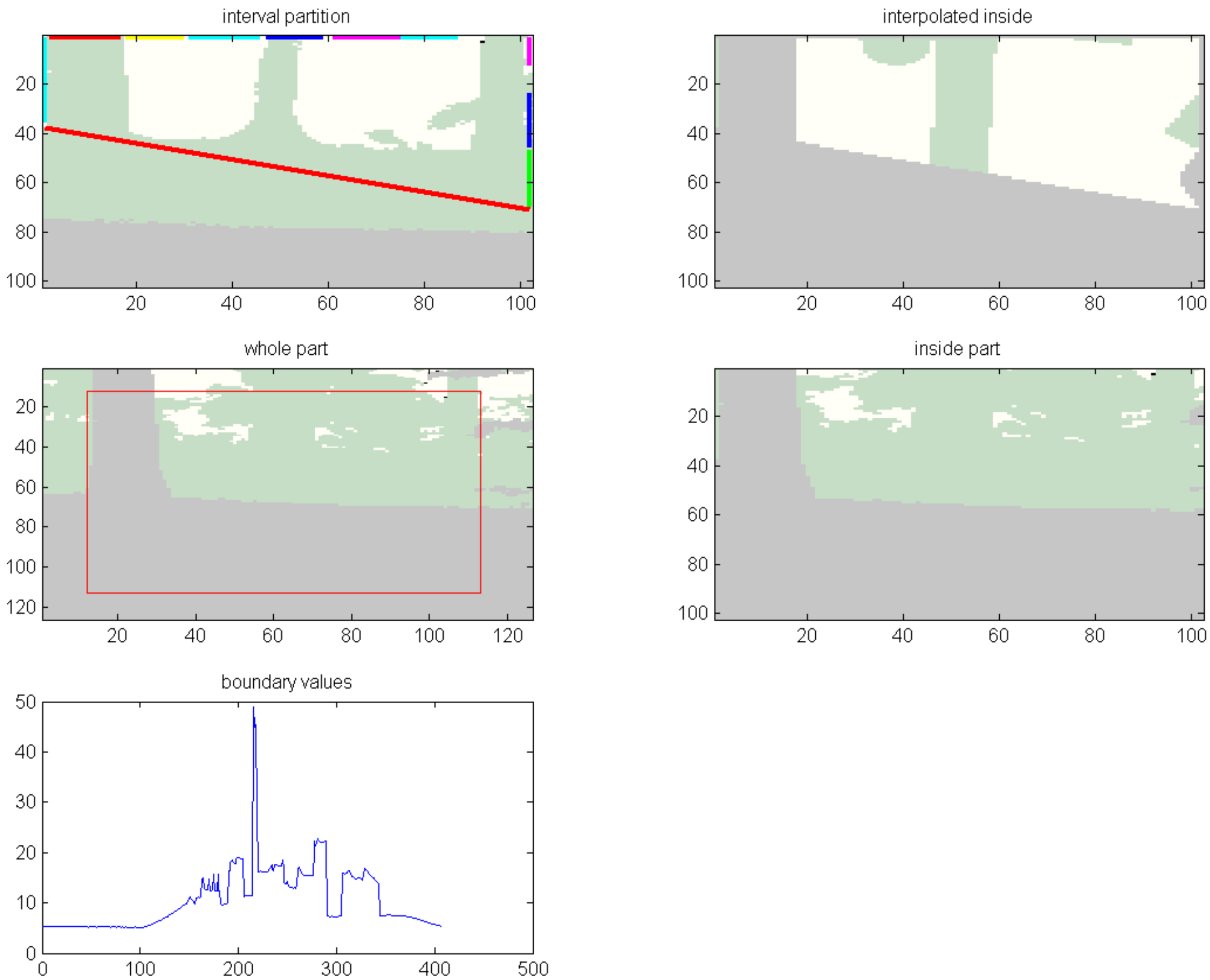

Figure 10

This scene has a big tree and two smaller ones further away. The inference is fairly good. And Figure 11

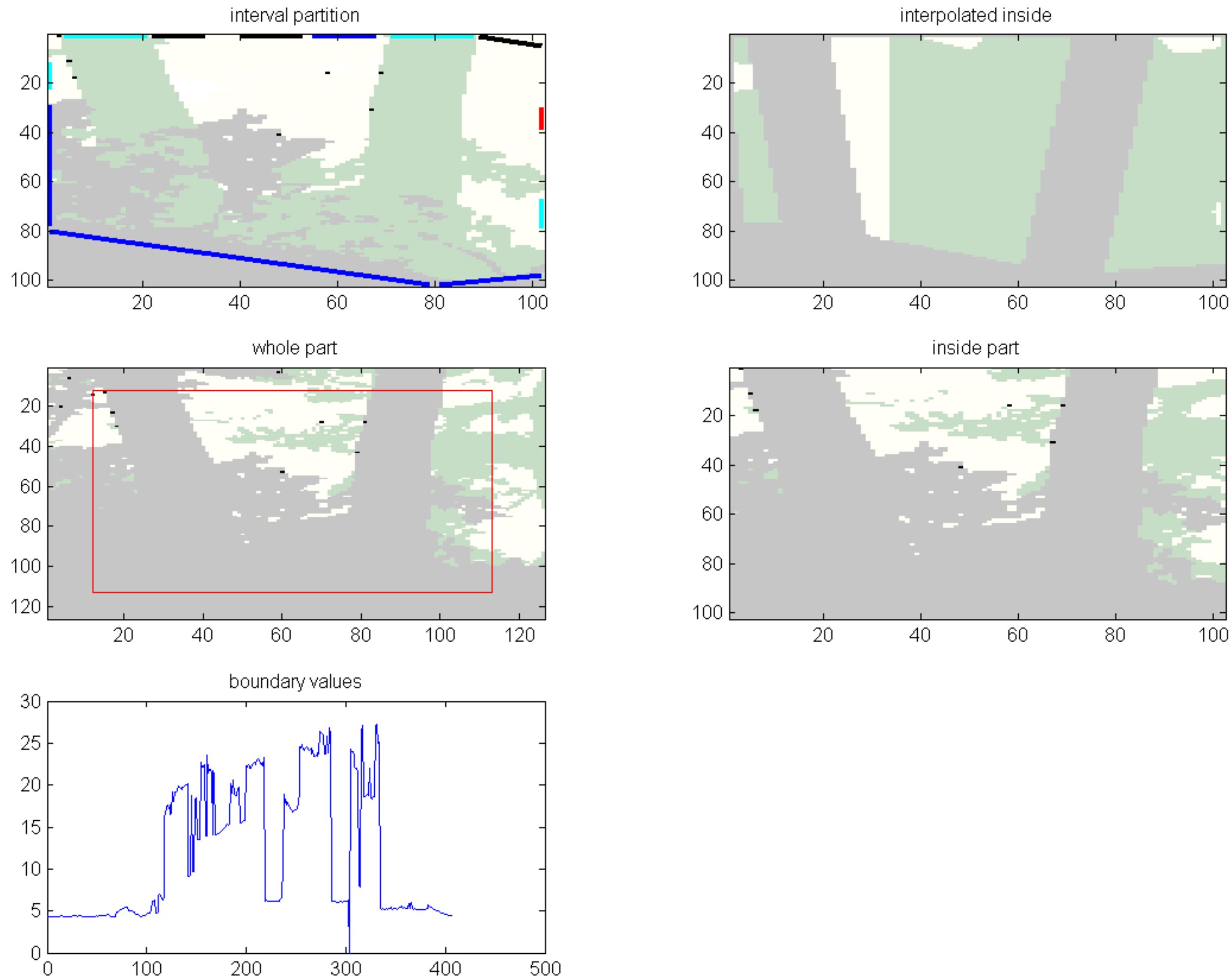

Figure 11

Also with two trees, but at about the same distance. Fine result but of course no algorithm could infer the bush between the trees from the frame data.

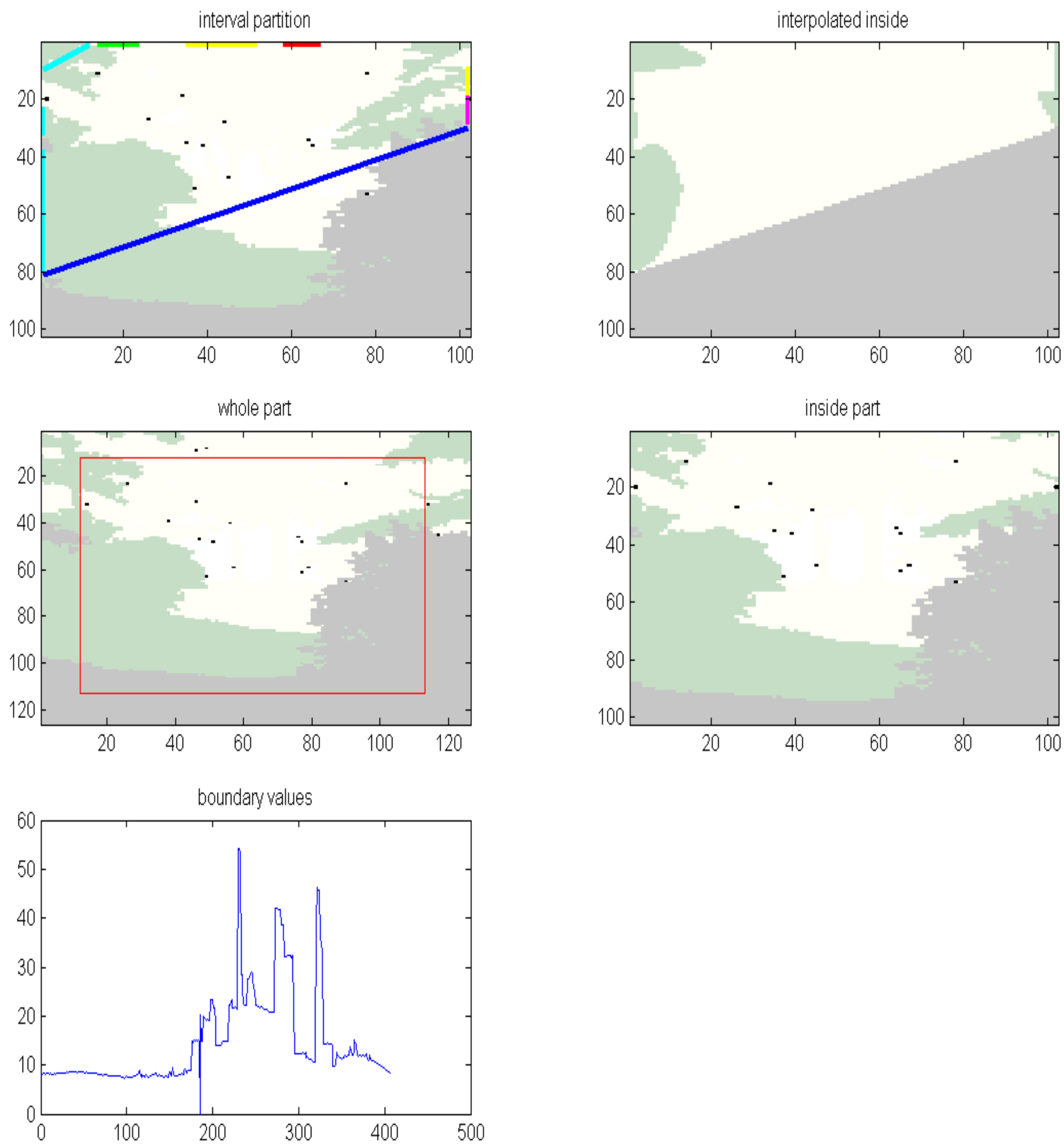

Figure 12

In Figure 12 we have big foliages to the right and to the left. The low boundary values below and to the right fool the algorithm to believe that the ground extends higher than it does. It makes a halfhearted attempt to continue the left foliage.

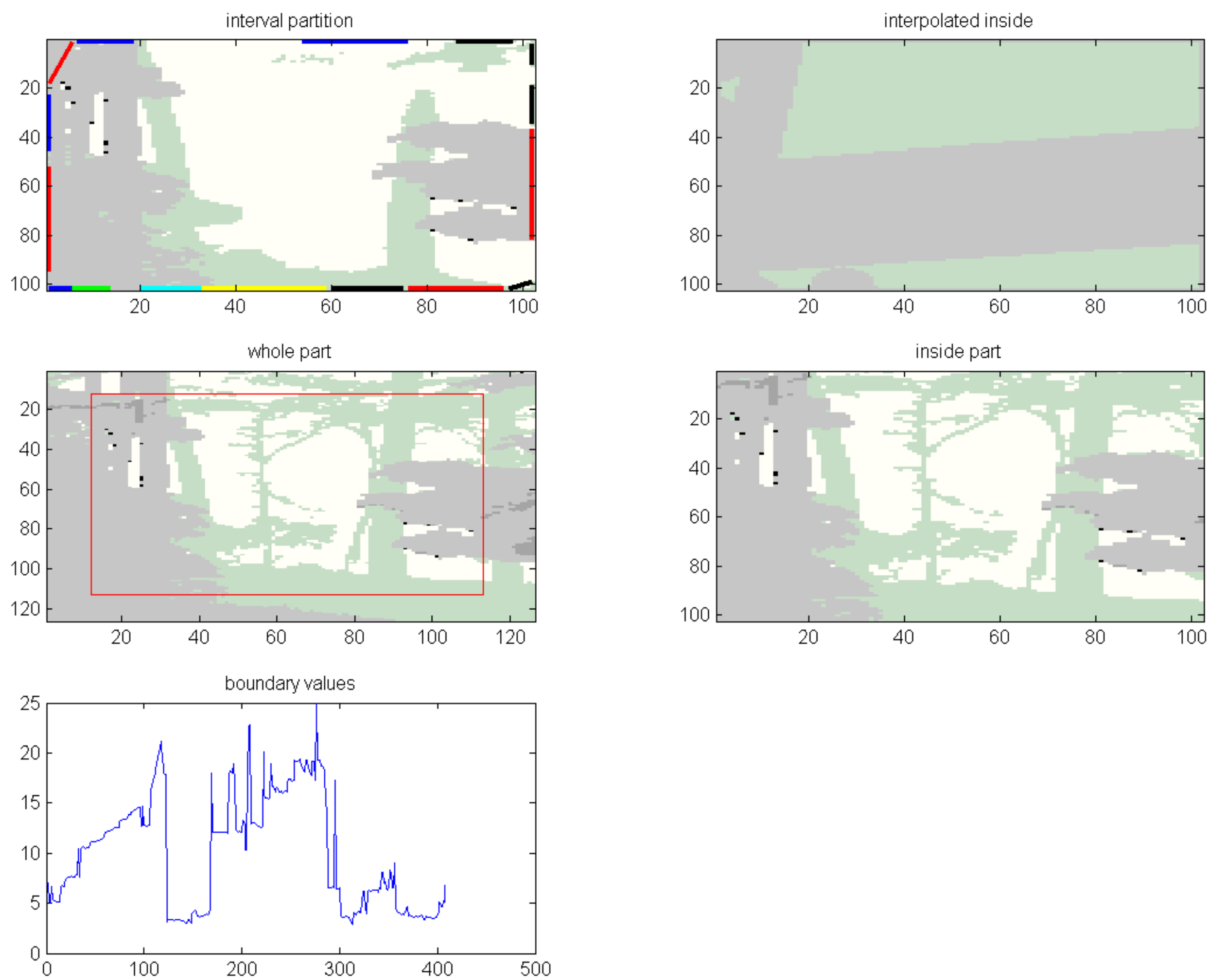

Figure 13

In Figure 13 we have similar effect: the right foliage has been continued much too far,

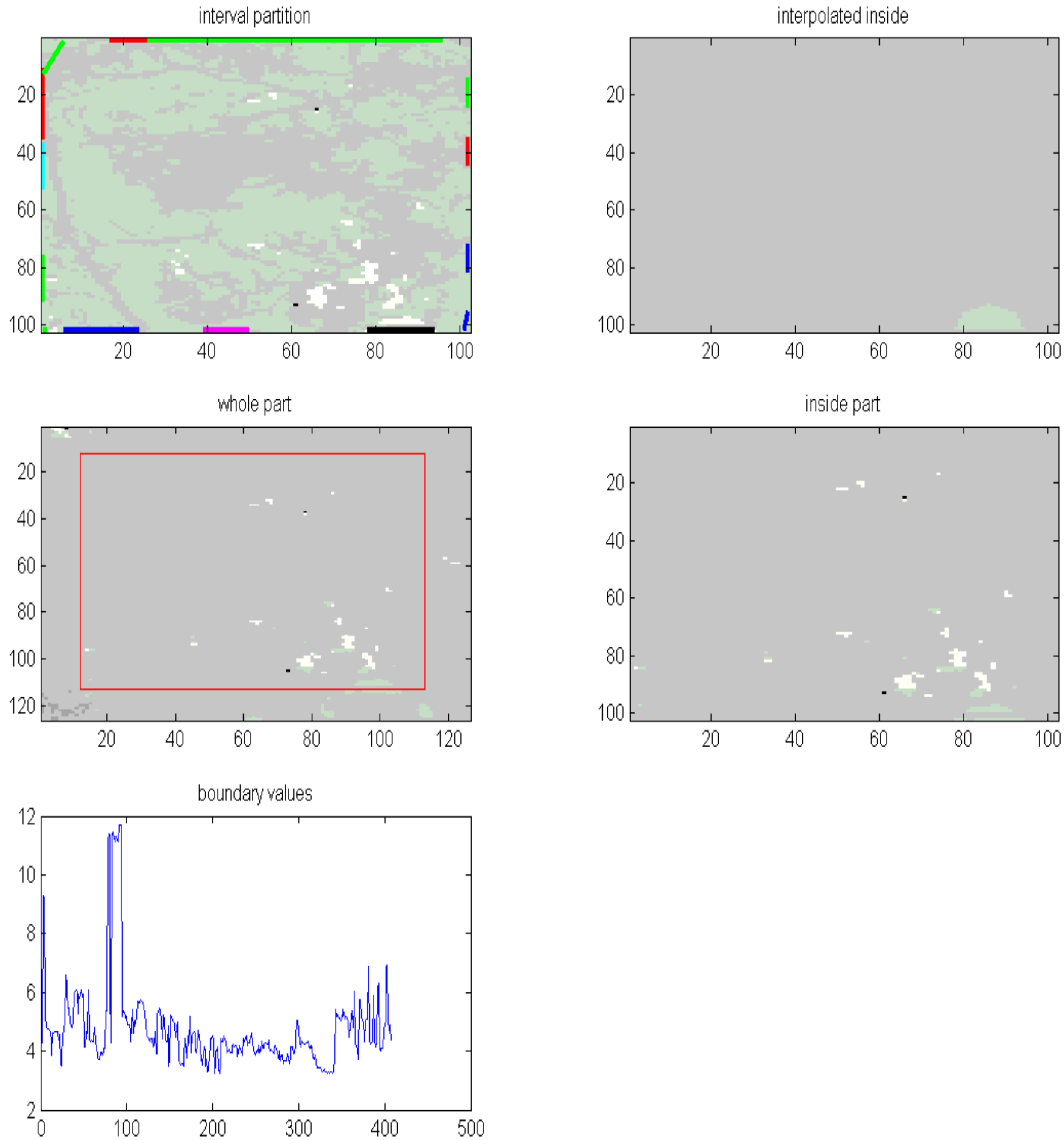

Figure 14

The algorithm is thoroughly confused in Figure 14. It just cannot understand such a complicated picture.

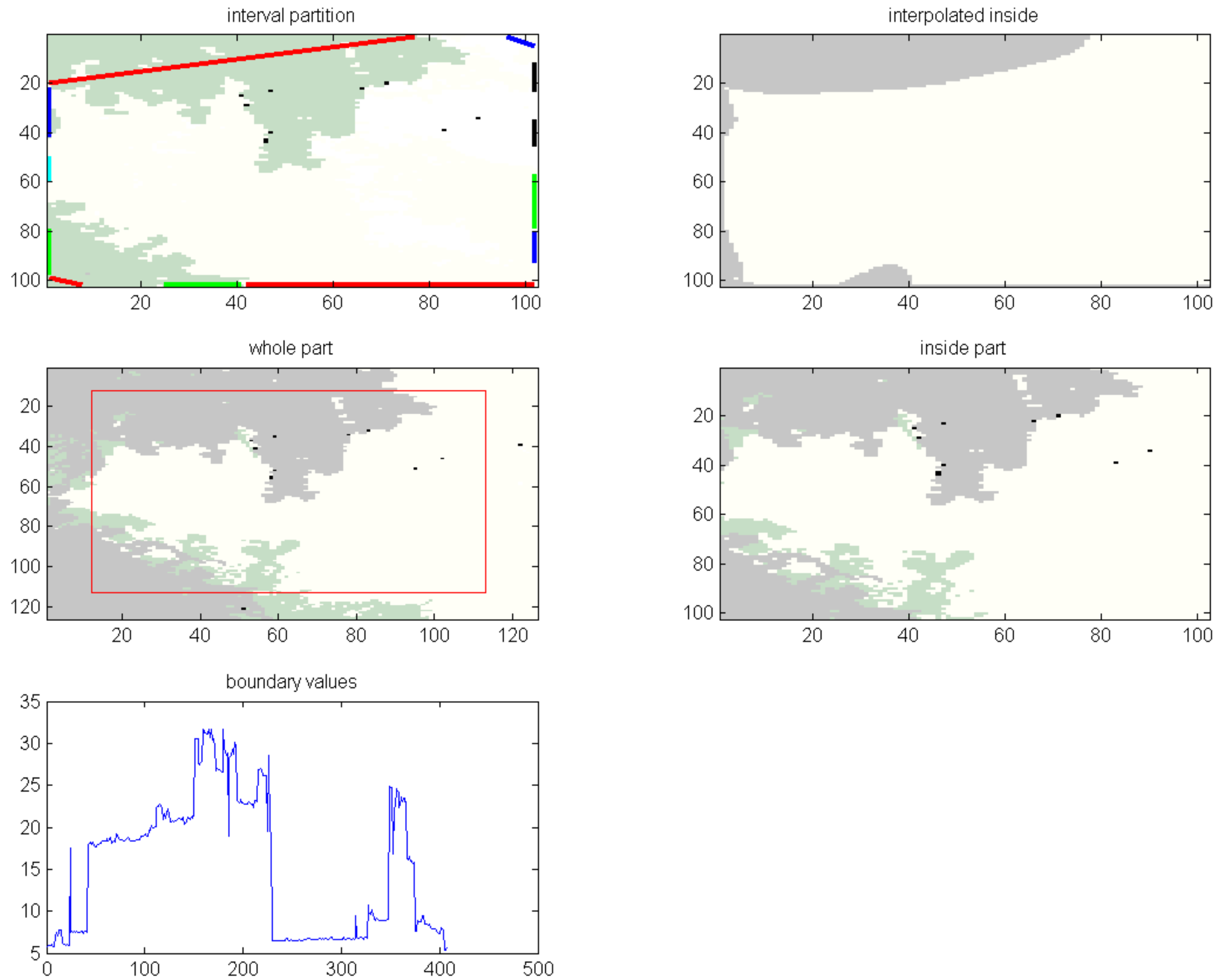

Figure 15

The algorithm suspects, rightly, the presence of a massive foliage in the upper left in

Figure 15.

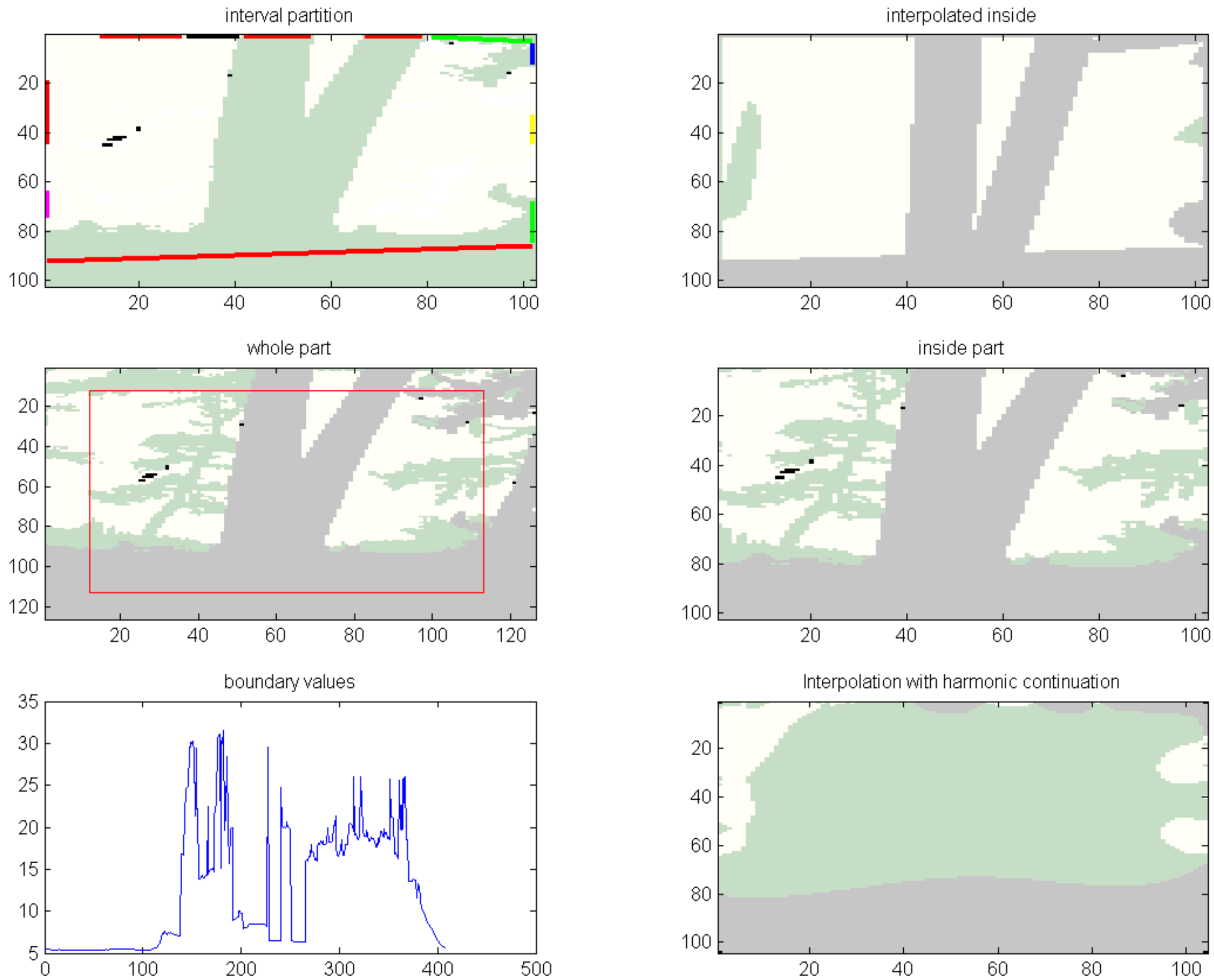

Figure 16

Almost perfect perception inference in Figure 16. For comparison the lower right panel we have shown the harmonic function interpolator,much worse.

5.7 How Did the Inference Algorithm Do?.

Not too badly. It could handle ground and trunk generators well, but the foliage presented greater problems. This is not surprising considering the greater variability of the foliage, but it could be handled better by improving the modules of the code for foliage recognition. Of course it can be questioned if any algorithm could perform really well for foliage.

It seems that the most vulnerable part of this module is the estimation of the direction tendencies. Often the inclinations are over estimated; perhaps this should be countered by systematicallyy avoiding large inclination estimates.

This also applies to the recognition of trunk generators: they are often too inclined left or right. Same remedy could be attempted.

A common error occurs when a trunk at the left or right boundary is partially visible together with ground, The algorithm then tends to think of this as a common ground element. It is not clear how this could be handled better.

When foliage appears it tends to interfere with the recognition of ground, trunk and sky generators. Possibly the program module could be modified to pay less attention to such foliage occurrences.

Although some of the inferences are poor, it seems that, considering the enormous variability of the scenery, that they are far superior to the non-specific ones; *the exploitation of trunk-foliage-sky-ground structure has helped a lot!* This approach has shown its strength and deserves further study and improvement, also for other types of inference.

6 Analysis.

To create a knowledge representation of forest scenes we first have to decide how detailed it should be. Should it involve individual trees and bushes? Must branches be specified? Should tree type be described in the representation, oak, maple, pine...? It all depends upon the goal we are set for using the representation.

If the goal is to discover man made Objects Of Interest, OOI, say vehicles or buildings, we may not need a very detailed descripton of the forest. On the other hand, if the purpose is to automate the collection of tree type statistics we should include tree type information in the knowledge representation. This is not just segmentation, it involves analysis and understanding of the content in the image.

Let us deal with the second of the two alternatives. With some arbitrariness we have chosen the following four generator indices:

A) α="ground" : ground surface in the foreground

B) α ="sky": sky background

C) α= "trunk": trunk element for individual trees

D) α= "foliage": close foliage

Now we must make the α-definitions precise. Since the TGM is 2D and lives in the image plane, the generators consist of areas in this plane. We therefore introduce *index operators* O^α mapping sub-images $I_X = \{I(x); x \in X\}$, where X is a subset, say a rectangle, in the image plane,

$$O^\alpha : I_X \mapsto \{TRUE, FALSE\} \tag{52}$$

In other words, the index operators are decision functions taking the value TRUE if we decide that the area X is (mainly) covered by a generator $g \in G^\alpha$. It may happen that a set X is classified as more than one α-value. We shall order the way we apply the index operators, one after the other, with the convention that a TRUE value overwrites previous truth values. We have used the order $ground, foliage, trunk, sky$.

In tabular form:

Generator Index Classes		
Generator Index	Informal Definition	Formal Definition
α = ground	Smooth Surface	{X:O^{ground}[I]=1}
α = foliage	Irregular Surface	{X:$O^{foliage}$ [I]=1}
α = trunk	Narrow Vertical Surface	{X: O^{trunk}[I]=1}
α = sky	Infinite Distance	{X: O^{sky}[I]=1}

6.1 Index operator for"ground"

A ground area is usually fairly flat and more or less horizontal except for the presence of boulders and low vegetation like bushes. We shall formalize this by asking that the gradient be small

$$O^{ground}[I(X)] = TRUE \leftrightarrow \|grad(I)(x)\| < c_1; x \in X \tag{53}$$

Of course the operator will fail in detecting sloping ground.

6.2 Index operator for"foliage"

For foliage on the other hand, the leaves give rise to considerable local variation but this variation is moderate as long as the leaves belong to the same tree. Hence we introduce

$$O^{foliage}[I(X)] = TRUE \leftrightarrow c_2 < \|grad(I)(x)\| < c_3; x \in X \tag{54}$$

6.3 Index operator for"trunk"

Trunk areas are narrow and vertical with small local variation compared to their immediate environment. In Figure 17 the narrow rectangle Xcenter should correspond to a part of a trunk, it is surrounded by two somewhat larger rectangles Xleft and Xright with some overlap. Compute the variances of the pixel values belonging to these three sub-images, call them $Var_{left}, Var_{center}, Var_{right}$ and define the index operator

$$O^{trunk}[I(X)] = TRUE \leftrightarrow Var_{left} > c_4 Var_{center} \; and$$

$$Var_{right} > c_4 Var_{center}; x \in X \tag{55}$$

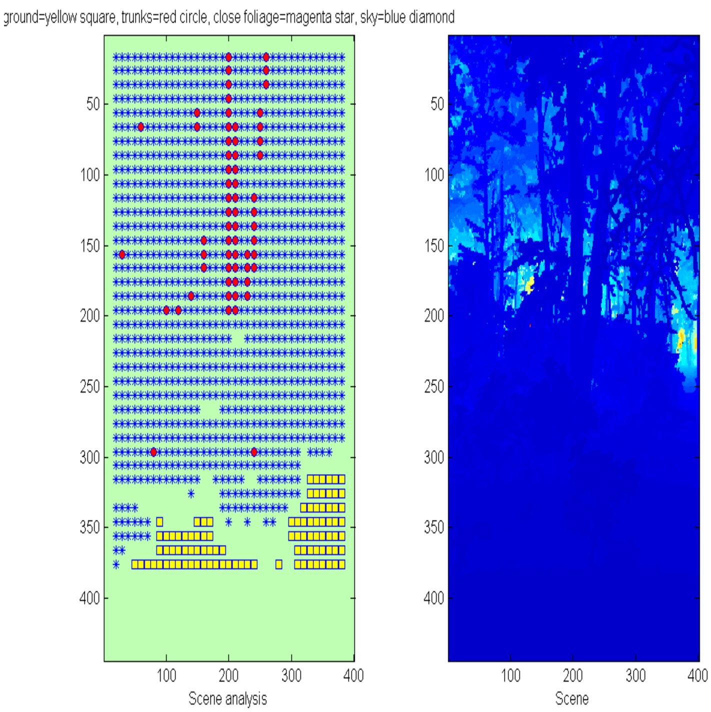

Figure 17

6.4 Index operator for"sky"

This is the easiest generator class to detect. Indeed, sky means infinite distance, far away, or rather the largest distance that the laser radar can register. For the camera used, this is coded as $I(x) = 0$. Hence we can simply define

$$O^{sky}[I(X)] = TRUE \leftrightarrow I(x) = 0; x \in X \tag{56}$$

6.5 Application to Range Images

. Applying this to laser radar images from Lee-Huang(2000): Brown Image Data Base, we get the following analysis with the meanings displayed graphically.

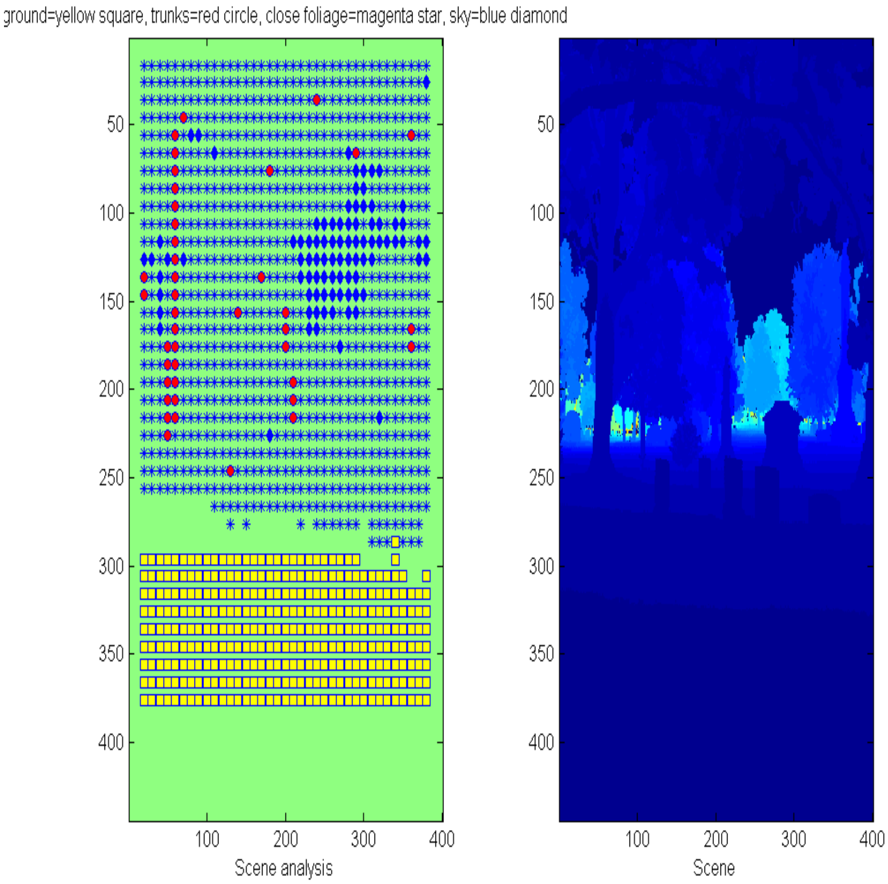

Figure 18

In Figure 18 one sees the occurrence of blue diamonds, "sky" and two trunks, some smaller trunk elements too, and a lot of foliage. At the bottom of the figure is the ground, separated from the foliage by pixels that could not be understood by the algorithm. Note that the tombstones in the observed scene have been interpreted as foliage: the present knowledge representation does not "understand" tomb stones. We shall study the understanding of such OOI's in Section 7.

In Figure 19 the dominating feature of the analysis is the occurrence of two big trunks. The ground is detected only in parts.

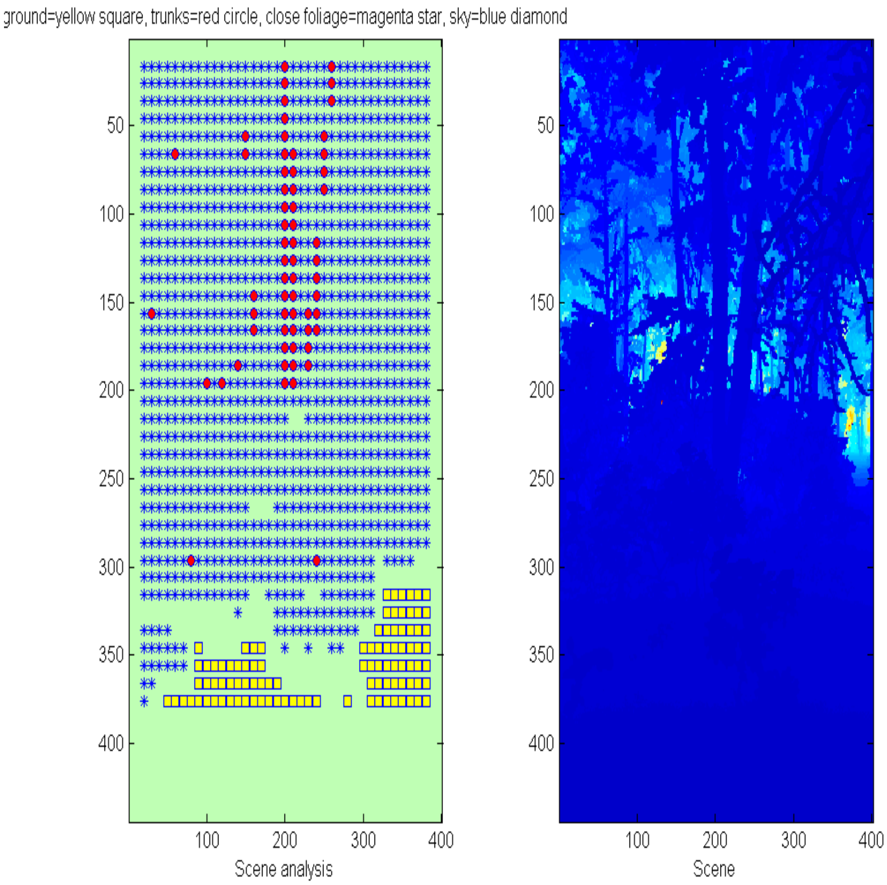

Figure 19

The analysis of Figure 20 also is dominated by a trunk. To much is interpreted as foliage.

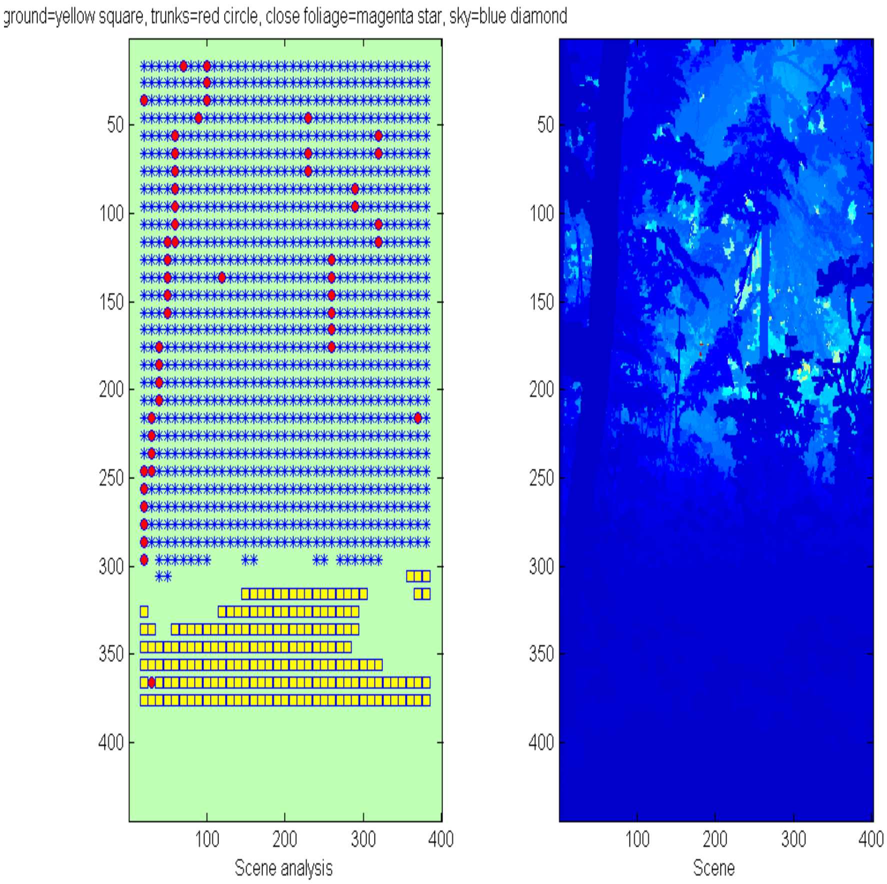

Figure 20

The analysis of Figure 21 has a sky element again.

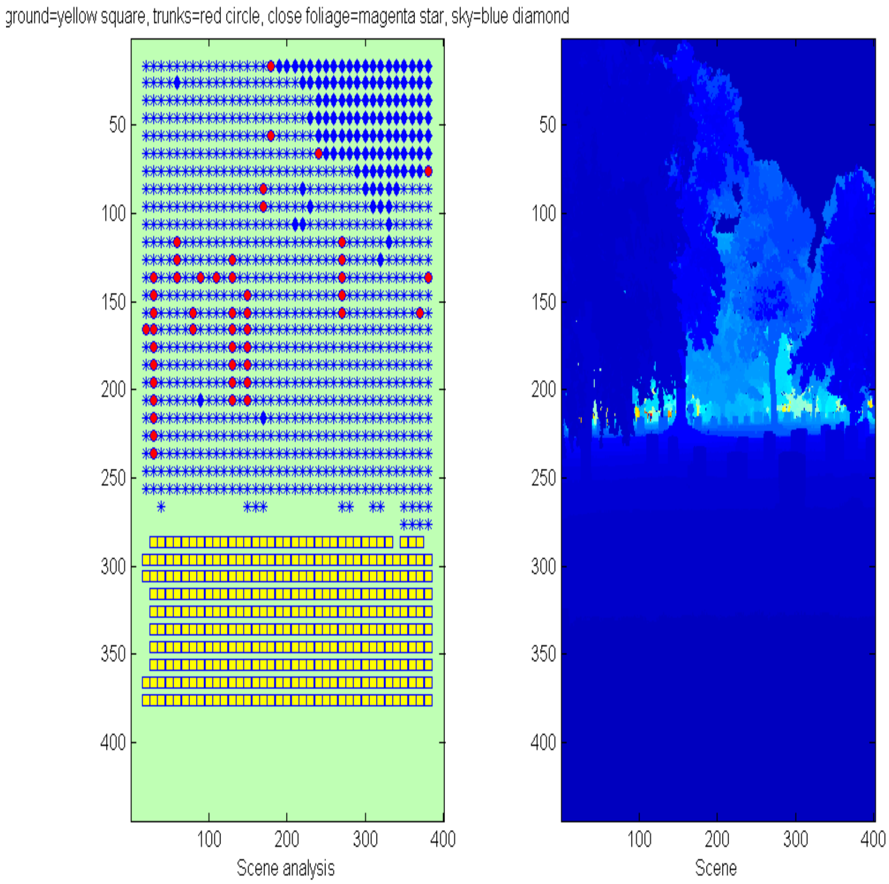

Figure 21

It seems that this automated image understanding performs fairly well considering the primitive form of the index operators that we have used. But it interprets to much as foliage and makes other mistakes as well .As we have argued repeatedly before, a major task in the creation of tools for the automated understanding of image ensembles is to *build specific, tailor made knowledge representations in pattern theoretic form for the image ensembles in question*. This is a real challenge, one that has been reached only partially here, and has successfully implemented only in some cases. In particular we would like to mention computational anatomy [], some microbiological image ensembles [], and certain object recognition scenes with movable rigid bodies [], but is a *sine qua non* for any serious attempt to automate scene understanding.

7 Recognition.

When we turn to the problem of automated recognition of OOI's against a background of clutter, say a forest scene, the role of the background changes. In the previous section the background, the forest, was of primary interest, but now it plays the role of nuisance parameter in statistical parlance. We are now not after an analysis into foliage, trunks..., but cannot neglect the clutter as irrelevant. It has long been known that the randomness of clutter is far from simple white noise. Further, it cannot be modelled as a stationary Gaussian process in the plane. Indeed the validity of the Bessel K hypothesis contradicts such a model from the very beginning. We shall use the TGM for representing the background clutter - the secondary element of the images - and the more detailed B3M for desribing the OOI's - the primary element. For the OOI we choose tanks; we happen to have available a template library in the form of CAD drawings with rotation angles equal to 0,5,10,15,20... degrees. Since we position them on the ground we will have $s_3 = 0$, so that this coordinate will be left out in the computations.

We shall assume the knowledge status

K =

knowledge element	element descriptor
k_1	one output from laser radar camera
k_2	one output from a FLIR
k_3	intelligence: tank of type T possibly present in scene
k_4	the shape of T given in CAD form

The image representation will take the form

$$I(u) = min\{I_{clutter}(u), Tsg^{OOI}(u); u \in \text{ } image\ plane\ U\} + e(u) \tag{57}$$

for the time being with only a single α-value for the OOI, and $e(\cdot)$ standing for the camera noise. For laser radars the noise level is low, perhaps white noise $e = N(0, \sigma^2), \sigma^2 << 1$. Further the clutter part of the image will be represented by the TGM as

$$I_{clutter} = \sum_{\nu} A_{\nu} g(u - u_{\nu}) \tag{58}$$

7.1 Detection.

A straightforward correlation detector will fail miserably, so that we have to modify it to take into account the workings of the range camera. In particular we should observe the role of the minimum operator in (???).

Denote the observed range image by $I^{\mathcal{D}}(u)$. If the noise level is low $I^{\mathcal{D}}(u) \approx I_{clutter}(u)$ for u-points not obscured by the OOI.

$$p(s|I^{\mathcal{D}}(\cdot)) \propto \pi(s)exp\{-\frac{1}{2\sigma^2}\|I^{\mathcal{D}} - min(I^{\mathcal{D}}, Tsg^{OOI})\|^2 \tag{59}$$

with a prior density $\pi(\cdot)$ on the similarity group S expressing the current information status.

We shall choose the attention field AF as a rectangle in the $(s1, s2)$-plane with the center at the *pointer* $(s1^{AF}, s2^{F})$ and with some width $width = (w1^{AF}, w2^{AF})$. It should be mentioned that we have used very different scaling for s_2 compared to s_1; changes in the former means much more than for the latter. Let us try a prior of the form

$$\pi(s) \propto exp[-\frac{(s1 - pointer(1))^2}{\sigma_1^2} - \frac{(s2 - ponter(2))^2}{\sigma_2^2}]; (s1, s2) \in AF; 0\ else \tag{60}$$

not depending on $s4$, so that the two first components are independent and Gausian when restricted to AF, while the fourth one, the rotation angle, is uniform on $\mathbf{T}$.

Now search for the MAP estimator. Starting at $s = pointer$ and using the Nelder-Mead algorithm, see Nelder,Mead (1965), for function minimization, we hope to get convergence to a local mimimum, hopefully close to the true s-value, at least if *width* is small enough. But the behavior is a bit puzzling. Sometimes it works well, sometimes not. Why is this so?

Of course, if most or all of the OOI is hidden by the clutter we cannot expect good inference. On the other hand, if only part of the OOI is hidden, the inference algorithm should work. This in contrast to methods that are not designed to take care of obscuration effects, for example simple correlation detectors.

7.2 Estimation of Location and Pose

. We now apply the algorithm to the range images in Huang,Lee (1999) and display the result by showing a subset of the image containing the OOI and also the *same* subset with the result of the algorithm. In each case the AF was chosen so that it covered the OOI but with a shift in the "pointer" away from the center of the object; this to correspond to mistakes in the knowledge status expressed through a prior on the AF due to limitations of the FLIR. We get Figure 22

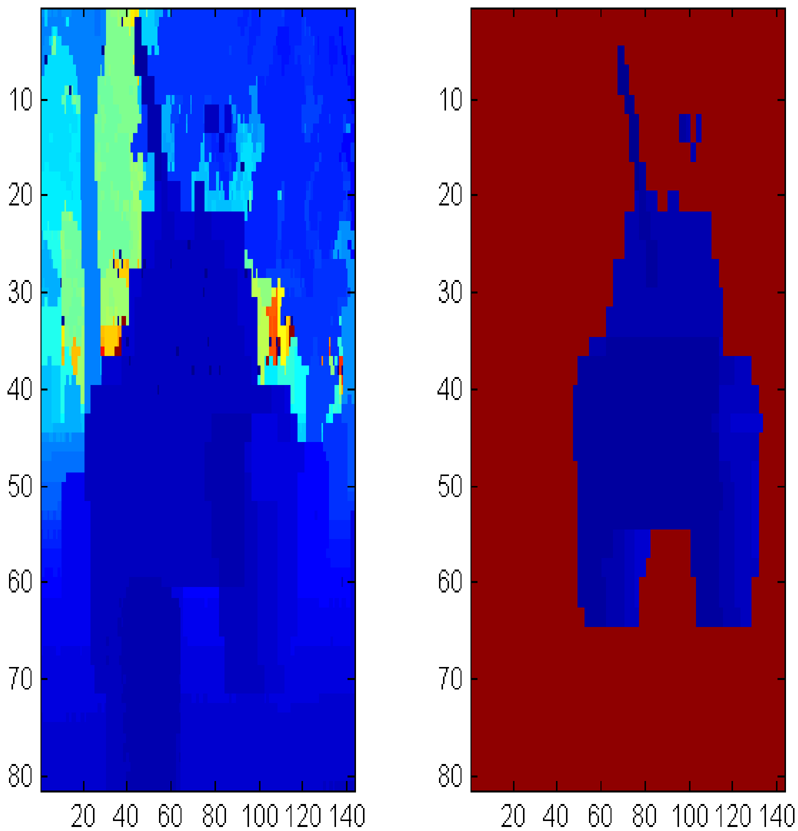

Figure 22

The inference looks good. Also for Figure 23

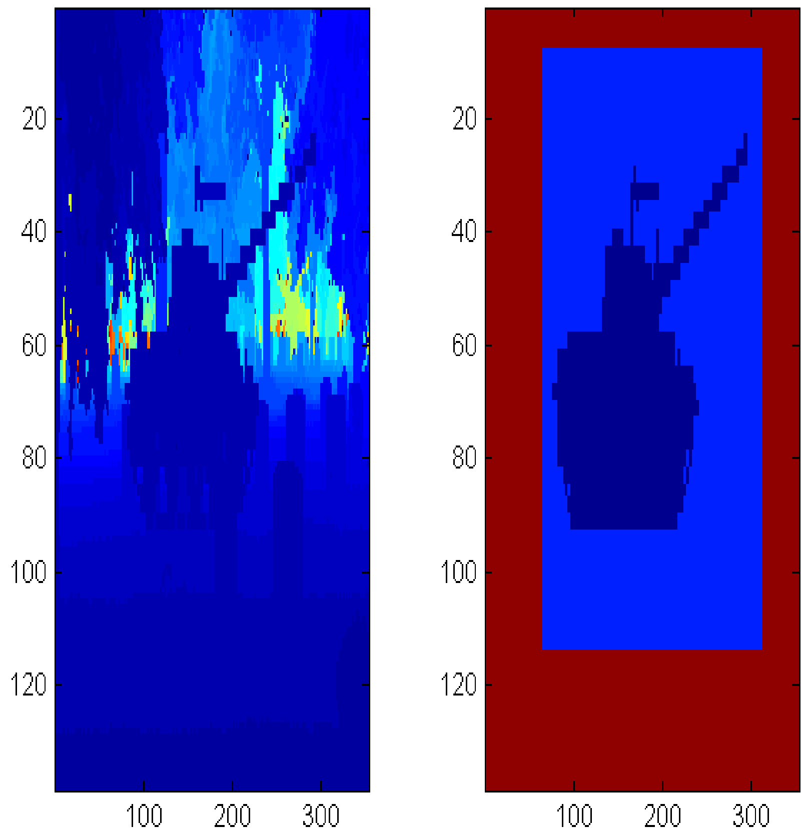

Figure 23

The same is true for the partly hidden OOI in Figure 24

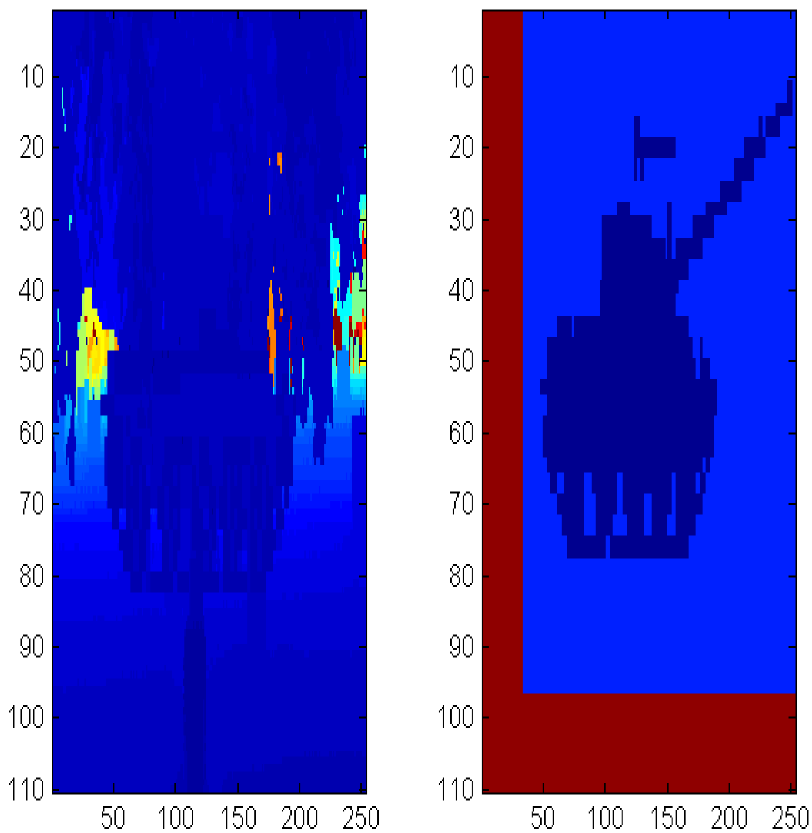

Figure 24
but the algorithm fails of course in the case of the wholly hidden OOI in Figure 25 and

does not discover the OOI

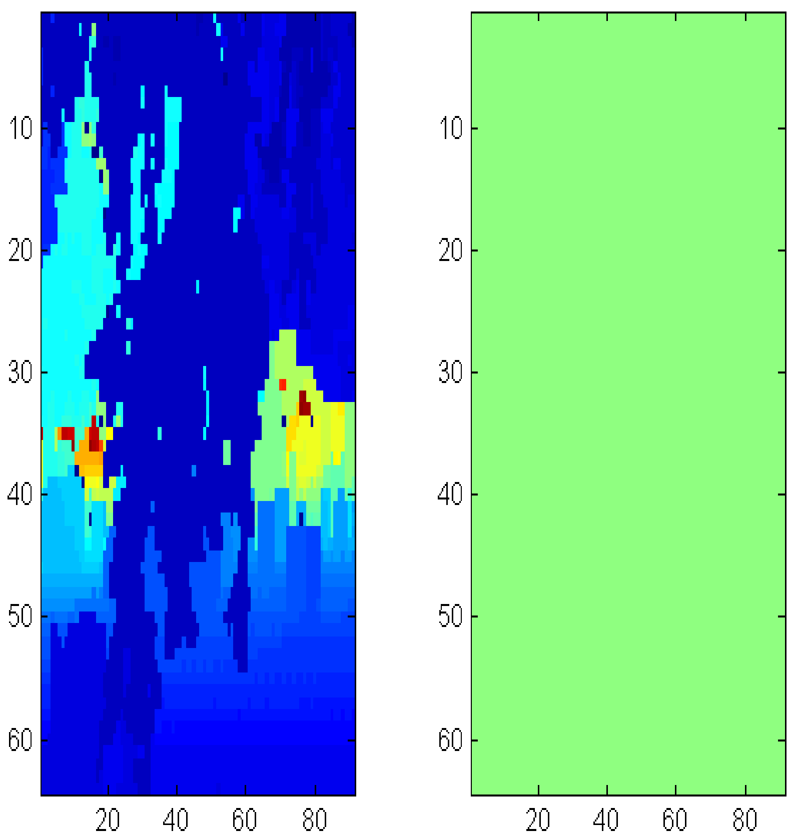

Figure 25

7.3 Trouble!

So far, so good. But sometimes the algorithm fails completely although the AF covers the true position of the OOI, this happens fairly often. Why is this? We had better look at the theoretical basis for this recognition algorithm.

The image matrix Tsg is of the same size as the observed image and with entries equal to $range_{max}$ outside of the projection of the transformed generator sg. We shall assume that for no values $s_1 \neq s_2$ is there a set $E, m(E) > 0$ in the visible part such that $(Ts^1g)(x) = (Ts^2g)(x), \forall x \in E$. This is a condition of non-self similarity. Actually, this is much stronger than needed for the following. For example, if the OOI is invariant with respect to some symmetry group, the statement in the theorem can easily be modified so that the new version is still true. As a matter of fact the OOI used in this study, a tank, has bilateral symmetry so that two rotation angles have the same visual effect.

Consider the estimate

$$I^*(u_1, u_2; s) = T(s_1, s_2, s_3, s_4) g_{temp}(u_1, u_2) \tag{61}$$

where we shall automatically set $s_3^* = 0$ as mentioned above. Recall that this estimate $I^*(u_1, u_2; s)$ takes values as functions of the matrix coordinates u_1, vertical directed downwards, and u_2, horizontal directed to the right. Its values outside the transformed template has been set to $range_{max}$. Find the visible part of OOI against a background $I_{clutter}$ as the support set

$$V(s) = \{(u_1, u_2) : I^*(u_1, u_2; s) \leq I_{clutter}\} \tag{62}$$

In particular the visible part of the OOI at the true value s^0 is $V(s^0)$ and we must have $m[V(s^0)] > 0$, positive Lebesgue area.

Now consider an alternative value s^1 of the group element, later to be assumed to be in a small neighborhood N of s^0, the notion of neighborhood to be qualified later. Consider the difference set

$$\Delta(s^1) = \{u : (Ts^1 g)(u) \leq (Ts^0 g)(u) \subset \mathbf{R}^2\} \tag{63}$$

meaning the part of the OOI for $s = s^1$ that can be seen in the presence of the OOI for $s = s^0$. We shall assume that

$$m[\Delta(s^1) \cap V(s^0)] > 0 \Rightarrow s^1 = s^0 \tag{64}$$

Or in words: inside the area of the OOI at the true position/orientation at least a bit of the OOI at the alternative position/orientation can be seen.

THEOREM: *If the observed image is the result of placing the transformed template against the backgound*

$$I^{\mathcal{D}} = min[I, Ts^0 g] \tag{65}$$

then, under the given conditions, the function $e(s), s \in N$

$$e(s) = \int_{u \in \mathbf{R}^2} [I^{\mathcal{D}}(u) - min[I^{\mathcal{D}}(u), (Tsg_{temp})(u)]]^2 du \tag{66}$$

has a local minimum at $s = s^0$.

PROOF: It is obvious that the non-negative function $e(\cdot)$ vanishes for $s = s^0$ since, with $I_{clutter}$ standing for the background image,

$$I^{\mathcal{D}} - min(I^{\mathcal{D}}, Tsg^0) = min(I_{clutter}, Ts^0 g) - min(min(I_{clutter}, Tsg^0), Tsg^0) = 0 \tag{67}$$

If, for an alternative value $s^1 \in N(s^0)$, the integral vanishes, we must have

$$I^{\mathcal{D}}(u) - min[I^{\mathcal{D}}(u)], (Ts^1 g_{temp})(u) = 0, a.e.u \in \mathbf{R}^2 \tag{68}$$

so that for $u \in V(s^0) = \{u : (Ts^0 g)(u) \leq I^{\mathcal{D}}(u)\}$

$$(Ts^0 g)(u) = min[(Ts^0 g)(u), (Ts^1 g)(u)] \tag{69}$$

or

$$(Ts^1 g)(u) \leq (Ts^0 g)(u) \tag{70}$$

Since visibility requires that $M[V(s^0)] > 0$ it follows from (???) that $s^1 = s^0$ and $s = s^0$ is indeed a local minimum in $N(s^0)$.

Now let us look more carefully at the local behavior of the $e(\cdot)$ function. Recall that with the background $I(\cdot)$ and the OOI resulting in $(Ts^0 g)(\cdot)$ in the image plane, and

$$e(s) = \int_{u:I^{\mathcal{D}}(u) > (Tsg)(u)} [I^{\mathcal{D}}(u) - (Tsg_{temp})(u)]]^2 du \tag{71}$$

so that

$$e(s) = \int_{u:I^{\mathcal{D}}(u) \geq (Tsg)(u); I(u) > (Ts^0 g)(u)} [(Ts^0 g)(u) - (Tsg_{temp})(u)]^2 du + \tag{72}$$

$$+ \int_{u:I^{\mathcal{D}}(u) \geq (Tsg)(u); I(u) < (Ts^0 g)(u)} [I(u) - (Tsg_{temp})(u)]^2 du + \tag{73}$$

Let $s = s^0 + \epsilon\eta$ where the 4-vector η has all its components equal to zero except for the kth one, $\eta_k = 1$ and $\epsilon \to 0$; the subscript k is one of the numbers 1,2,3,4. Consider the two integrals, I_1, in (?), and I_2 in(??). We can write

$$I_1 = \int_{u:(Ts^0 g)(u)(u) \geq (Tsg)(u); I(u) > (Ts^0 g)(u)} [(Ts^0 g)(u) - (Tsg_{temp})(u)]^2 du \tag{74}$$

Also

$$I_2 = \int_{u:(Tsg)(u) < I(u) < (Ts^0 g)(u); I(u) < (Ts^0 g)(u)} [I(u) - (Tsg_{temp})(u)]^2 du \tag{75}$$

Begin with I_1. If $u \in interior\{support[(Ts^0 g)(\cdot)]\}$ we have for small ϵ

$$(Tsg)(u) - (Ts^0 g_{temp})(u) = \epsilon \times constant + O(\epsilon^2) \tag{76}$$

so that the contribution to the integral I_1 from the interior is asymptotically propotional to ϵ. But the asymptotic contribution from a thin band around $\partial[(Ts^0g)(\cdot)]$ of I_1 depends upon the sign of ϵ. Indeed, if we are in $\{u : (Ts^0g(u) \geq (Tsg)(u)\}$ the contribution will again be proportional to ϵ asymtotitcally, generically with a positive proportionality constant. On the other hand for the opposite sign of ϵ the domain of integration reduces to the empty set. Hence, generically, the partial derivative of I_1 w.r.t. *epsilon* will not exist: the left and right derivatives will be different. We can discuss I_2 in the same way.

Because of this lack of differentiability we shall avoid search algorithms to get s^0 that assume smoothness. Instead we shall use the following primitive algorithm. We shall search cyclically over the coordinates $k = 1, 2, 3, 4$. In the tth iteration say that we have reached the value $s(t)$. Then, for each k, we shall consider the three values

$$e[s_k(t) - ds_k(t)], e[s_k(t)], e[s_k(t) + ds_k(t)] \tag{77}$$

where we will choose

$$ds_k(t) = f_k(t) \downarrow 0 \tag{78}$$

To get convergence to any true s^0 value we shall ask that

$$\sum_{t=1}^{\infty} f_k(t) = +\infty \tag{79}$$

In the software we used $f_k(t) = 1/\sqrt{(}t)$. Then pick $s_k(t+1)$ as the one of the three s_k values with the smallest $e[s(t)]$ value.

But this is not enough. It will be instructive to look at the behavior of the e function.

Typically it looks like the one in Figure 26

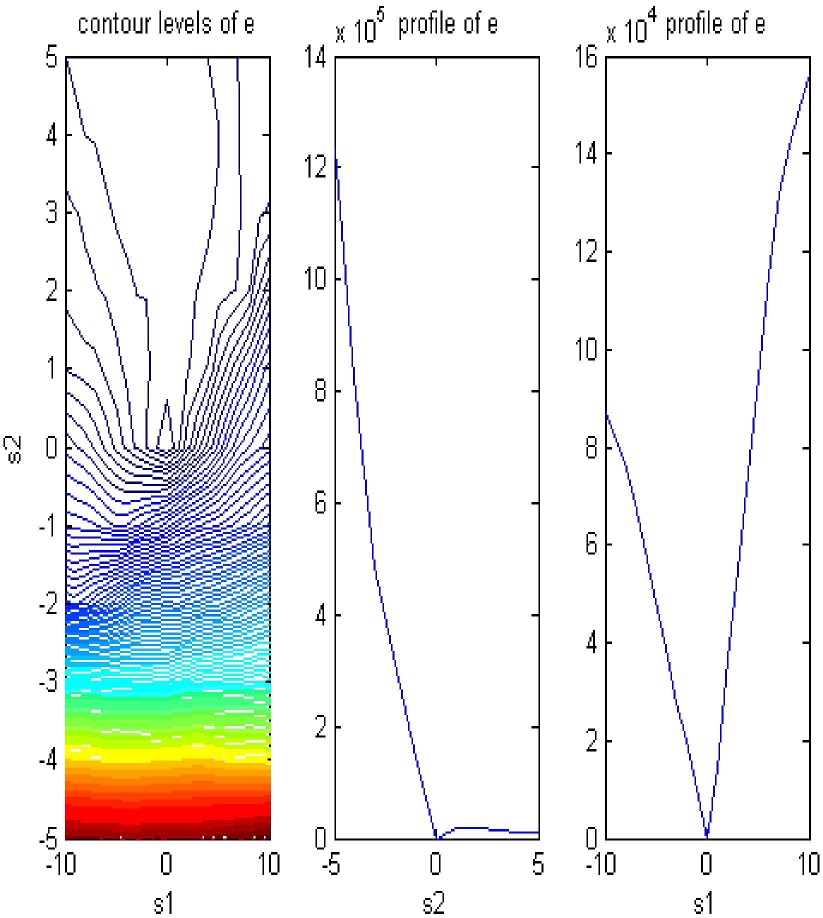

Figure 26

in which the left panel shows contour lines of the e-function with the minimum $e = 0$ at the value $s = (0, 0)$. Note how close the contours are in the halfplane H in contrast to in the other halplane, indicating rapid decrease toward the minimum. The middle panel show a profile $e(0, s_2)$, non-differentiable at $(s_2 = 0)$, which agrees with theory. The right panel shows the profile $e(s_1, 0)$, also with a jump in the derivative. Of course, for real data we can not expect to get the minimum exactly equal to zero due to measurement and numerical noise.

But Figure 26 also teaches us something else that we have hinted at earlier. The right panle, showing the dependence upon s_1, the side ways coordinate, increases fast as we move away from the minimum e-value, both to the right and the left: the minimum is well defined. In the middle panel, however, showing the dependence on s_2, the depth in the image, the left branch mincreases as we move away from the minimum. The right branch on the other hand, increases first, slowly, but then has a small local maximum and then decreases. This explains why the straightforward appliction of the algorithm for minimizing $e(\cdot)$ in the previous section did not always work: the occurrence of the *min* operation in (????), typical for range cameras. If we let s_2, the distance away from the camera, be large, the OOI will eventually be hidden behind the clutter, so that $ID - min(ID, Tsg^{OOI}) \equiv 0$, and only the π factor in (???) will play a role in the minimization; the information in the observed image will have been wasted, and we will get misleading inferences. It follows that if we use straight minimization estimation constraints, *the estimator is not consistent.* Instead we should use only a neighborhood

$$N(s^0) \subset H; H = \{s : s_2 \leq s_2^0\} \tag{80}$$

constraining the neighborhood to be in the halfplane H. In order to make the inference algorithm work we should therefore, *introduce a bias in the choice of the pointer* favoring smaller values of $s2$ in the AF. But we do not know where the minimum is located in the AF given by the FLIR. Let us therefore use a conservative search strategy, starting the search in the midpoint of the lower boundary of the attention field. If we do this we will find that the inferences are often quite good.

7.4 Recognition With Biased Starting Point.

Returning to the recognition algorithm but now constrained to a half plane, we shall illustrate the convergence of the algorithm by showing the successive values as curves, *saccadic (correct spelling in figures) paths in the group.* Another difference is that we shall use larger AF's than in Section 7. Of course we do not know where the halfplane H is actually situated in AF, so we shall use a conservative strategy, starting the minimum search at the midpoint ofthe AF's lower side. We use a very crude minimization algorithm here; better behavior can be expected with some industrial strength algorithm. In Figure 27 we see the siccadic path settling down around $s = (30, 1.75, 0, 270)$, indicate by a small circle, while the true vlue is $s = (30, 2, 0, 65)$, where the angle 65 has the supplement 295. Good behavior

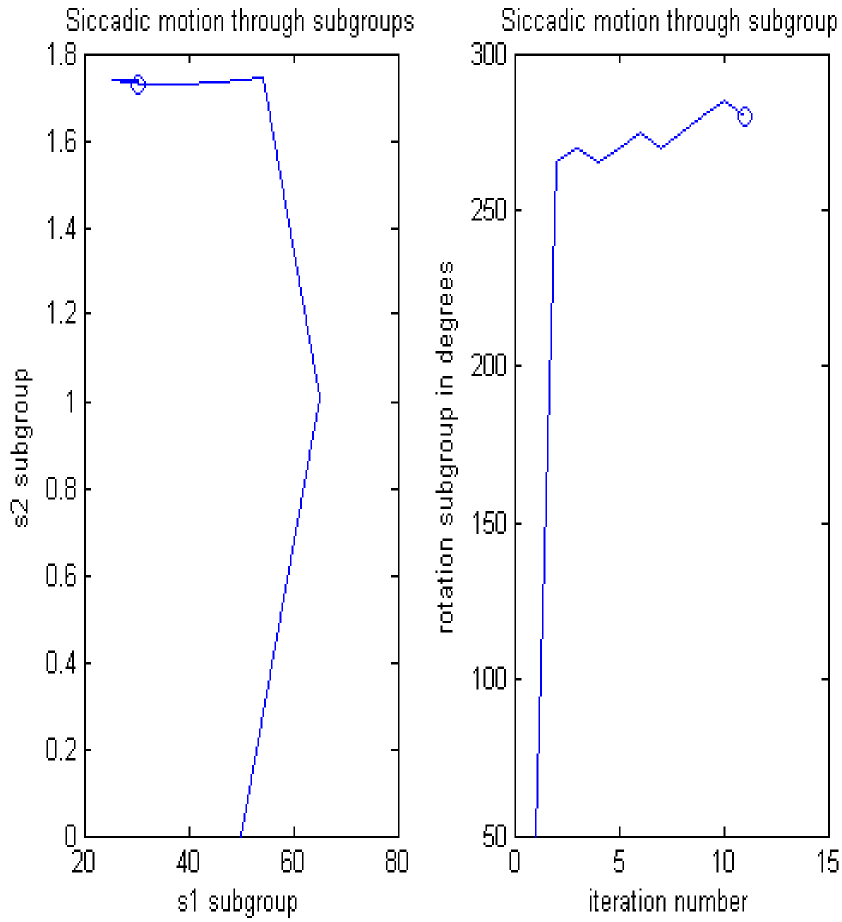

Figure 27

On the other hand in Figure 28 the saccadic limit $s = (32, 3.7, 0, 90)$ is far away from the true value $s = (-20, 4, 0, 60)$. The reason is that the distance of the OOI from the camera is greater (recall that we have used very different scaling of s_1 and s_2), so that it appears as only a small part of the observed image; little evidence for the inference.

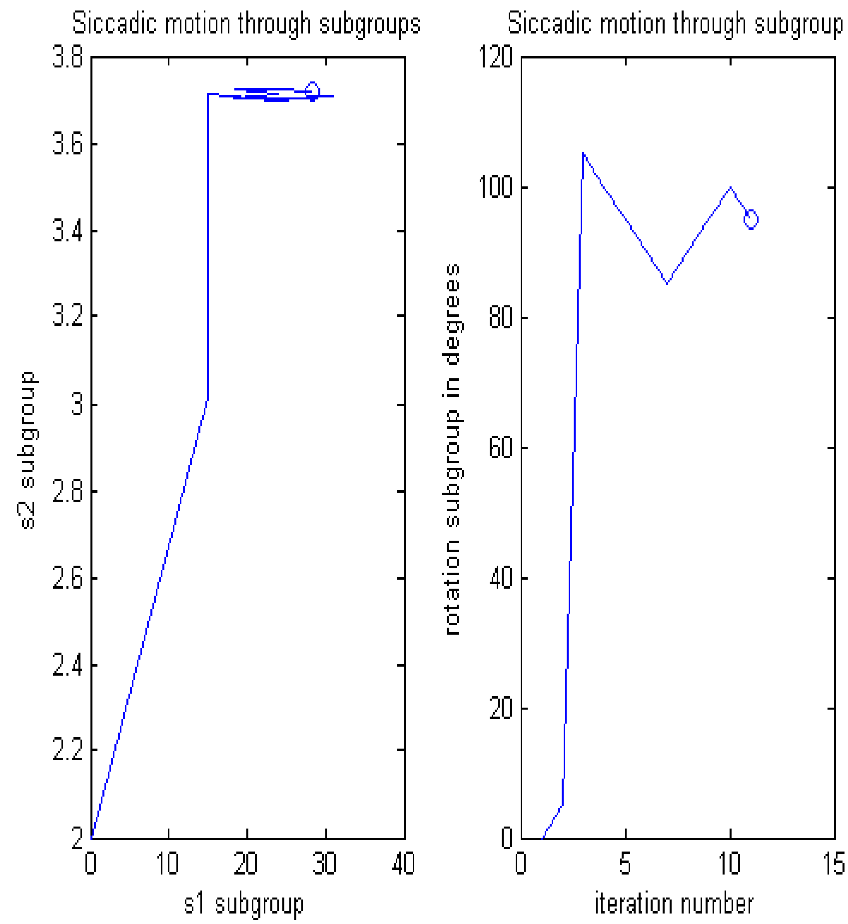

Figure 28

The performance is fairly poor in Figure 29 where the saccadic limit is about $(5, -.3, 0, 90$ to be compared with the true value of location/pose $(0, 0, 0, 0, 200)$, at least for the orientation estimate.

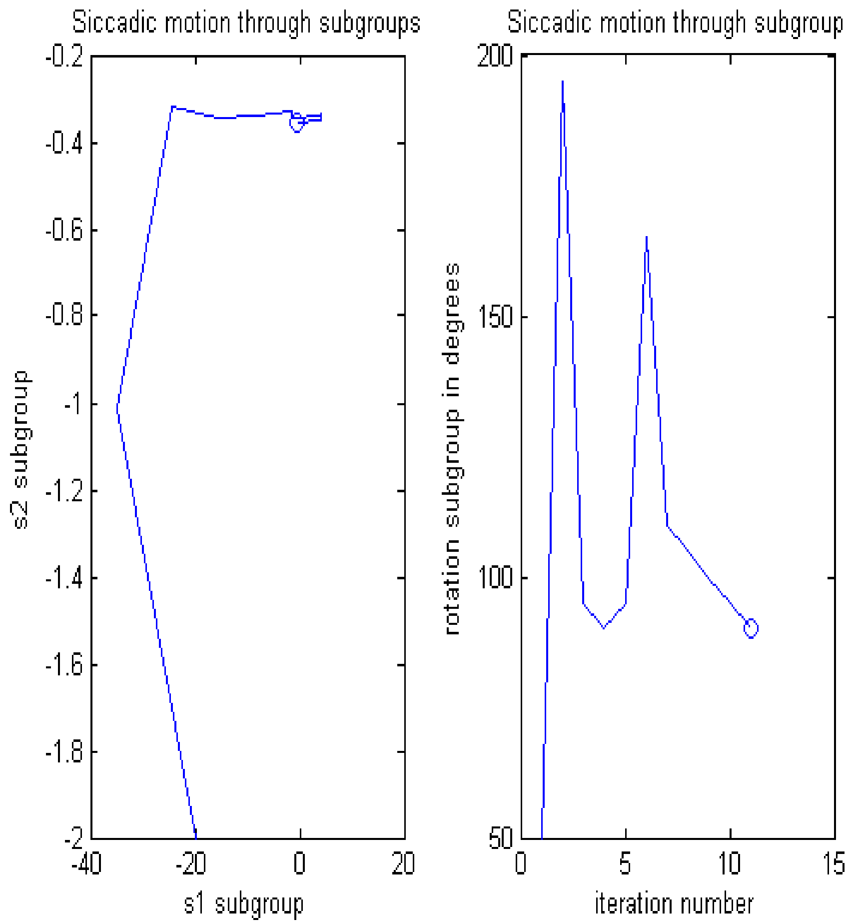

Figure 29

In Figure 30 the agreement is very good. Estimated parameters $(102, -1.3, 0, 90)$ compared with the true values $(100, -1, 0, 100)$

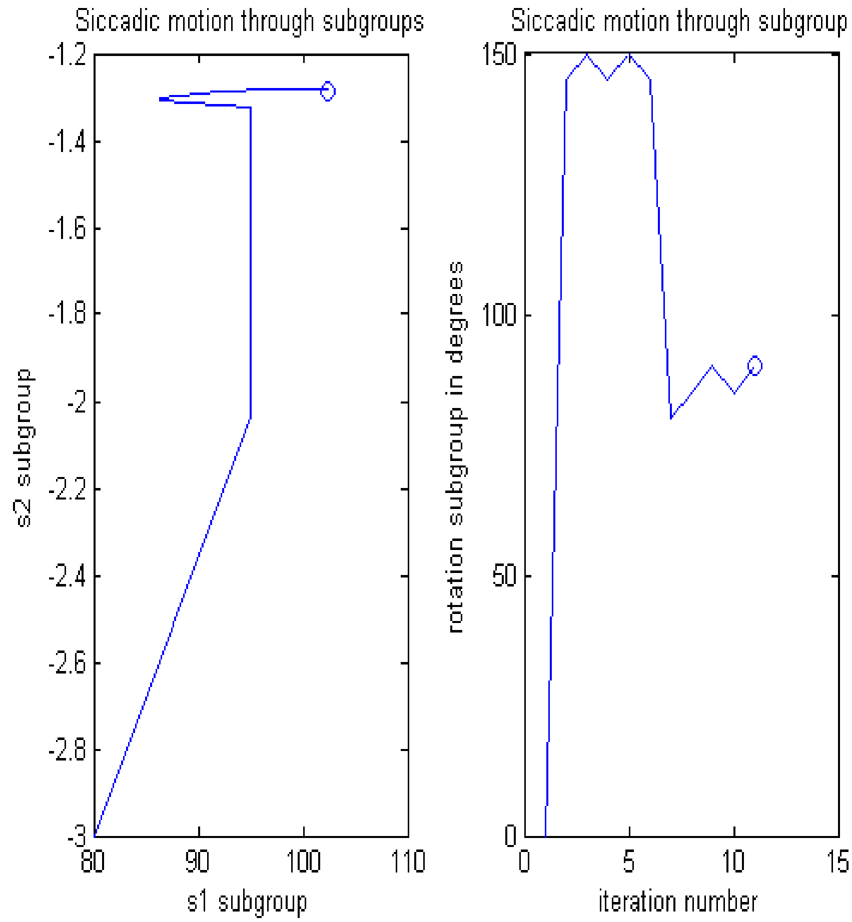

Figure 30

A good deal of the tank is hidden in Figure 31. Considering this the result of the detection estimation is perhaps not too bad, $(233, -.3, 0, 130)$ instead of the true $(300, 0, 0, 0, 90)$ but the location estimate is far too much to the left. The hiding trees to the left are fairly close in distance from the tank which of course makes estimation of s_1 difficult. In principle this comment is valid in general.

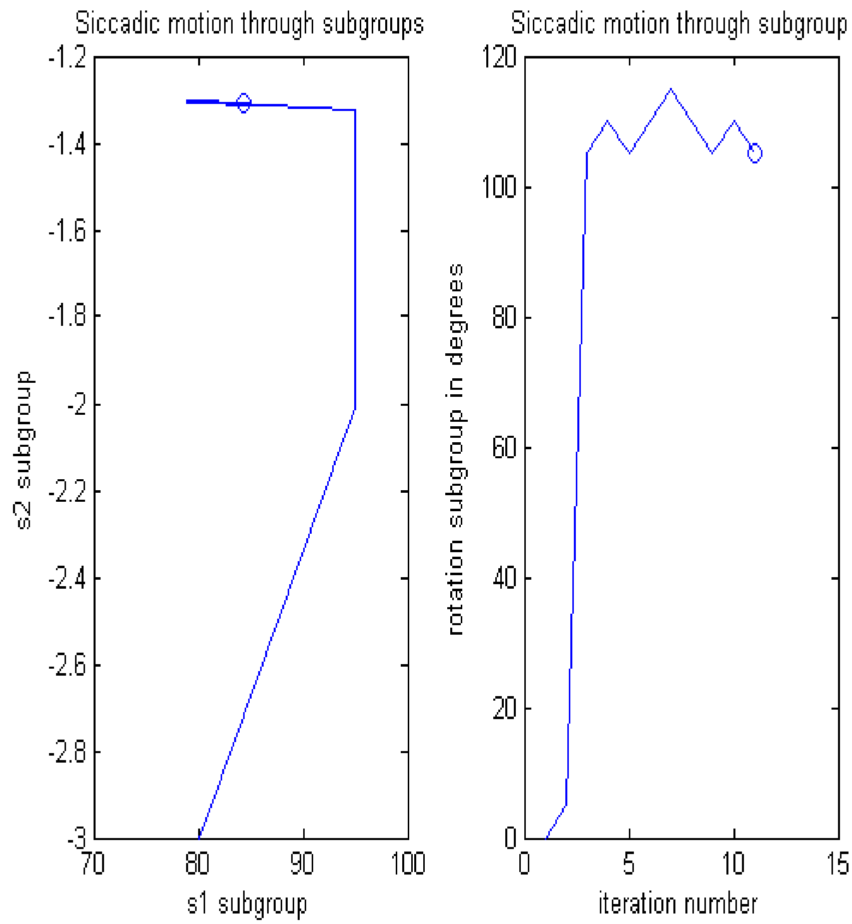

Figure 31

The observed $I\mathcal{D}$ is seen in Figure 32 with the tank to the right and further away from the trees, partly hidden up in the trees (sic!).

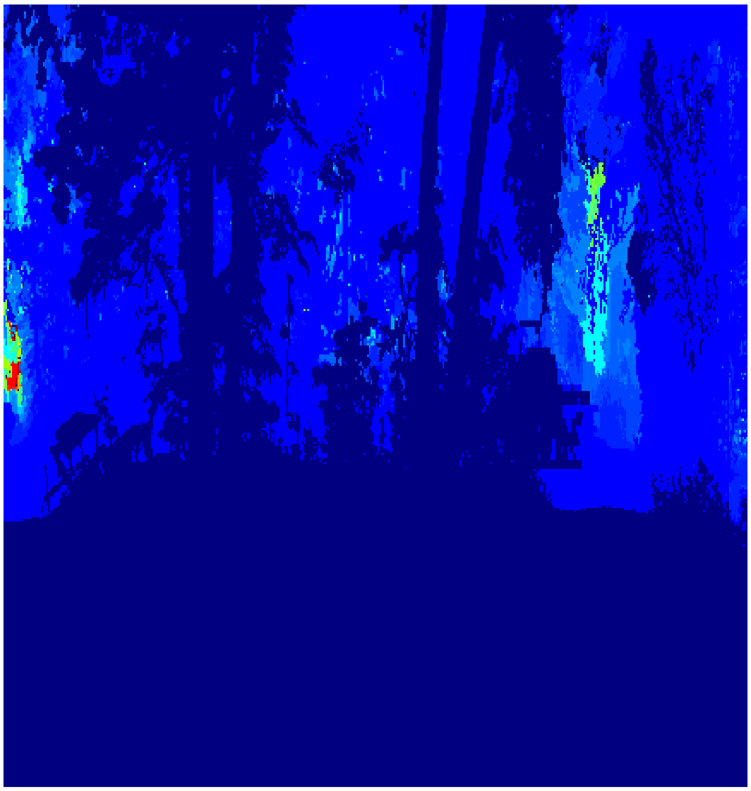

Figure 32

Here we have used biased selection of the initial poin in the half space H; starting in H^c often results in failure as predicted by the proof of the theorem.

8 Recognition

. Once the OOI has been found in terms of location and pose the way we have described, it should be possible to implement a scheme of understanding by running the algorithm for several different template libraries, to see which type of OOI we have in the observed image. We should use several α-values, see equation (10). Here we can only touch on the question, leaving a full treatment till later.

We shall use two template libraries, $G^{(1)}$: small tanks, and $G^{(2)}$:big tanks. Actually $G^(1)$ is the library we have used in the previous section. Run the estimation algorithm for some clutter background $I_{clutter}$ and with templates from $G^{(1)}$ and $G^{(2)}$ respectively, and note the resulting minimum values of $e(\cdot)$. We got for the true α-value the minimum $e_{min}^{(1)} = 3.55 \times 10^4$ while the alternative value was $e_{min}^{(2)} = 6.34 \times 10^5$, a drastic difference in favor of the true hyposthesis.

We do not know to what extent this difference holds in general. Nor have we studied how to calculate the cutoff value analytically to separate two or several α-hypotheses.

9 What Have We Learnt ?

This study has taught us the following.

(i) To handle the joint prior measures for natural scenes needed for inference it is neccessary to use clutter specifications tailored to the particular type of background that is expected in the knowledge status.

(ii) Detection, estimation of location and pose for (partially hidden) OOI's against a cluttered background can be organized as a minimum $e(\cdot)$ problem.

(iii) It appears possible to achieve recognition using the technology of (ii).

10 Bibliography.

A. J. Bell and T. J. Sejnowski. The independent components" of natural scenes are edge filters". Vision Research, 37(23):33273338, 1997.

Z. Chi, Probability Models for Complex Systems, PhD thesis, Division of Applied Math., Brown Univ., 1998.

U. Grenander. General Pattern Theory. Oxford University Press, 1993.

U. Grenander: Clutter 5,6,1999a,b, www.dam.brown.edu/ptg.

U. Grenander, M.I. Miller, and P. Tyagi, Transported Generator Clutter Models, Monograph of Center for Imaging Sciences. Johns Hopkins Univ. 1999.

U. Grenander and A. Srivastava. Probability models for clutter in natural images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(4), April 2001.

U. Grenander, A. Srivastava, and M. I. Miller. Asymptotic performance analysis of bayesian object recognition. IEEE Transactions of Information Theory, 46(4):16581666, 2000.

U. Grenander and M. I. Miller (1994): Representation of Knowledge In Complex Systems (with discussion section), Journal of the Royal Statitical Society, vol. 56, No. 4, pp. 569-603.

U. Grenander and M. I. Miller (1996): Computational Anatomy: An Emerging Discipline, feature article, Statistical Computing and Graphics Newsletter, vol. 7, no. 3, pp. 3-8.

J. Huang and D. Mumford, Statistics of Natural Images and Models, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 541-547, 1999.

J. Huang and A.B. Lee : Range Image Data Base, Metric Pattern Theory Collaborative, Brown University, Providence RI,1999, www.dam.brown.edu/ptg

A.B. Lee, D. Mumford, and J. Huang, Occlusion Models for Natural Images: A Statistical Study of Scale-Invariant Dead Leaves Model, Int. J. Computer Vision, 2000.

J. Huang and D. Mumford: Statistics of Natural Images and Models". Proc. IEEE Conference Computer Vision and Pattern Recognition, 541-547, 1999

Ann. B. Lee: Statistics, Models and Learning in BCM Theory of a Natural Visual Environment, Ph.D. thesis, Brown University, 2002.

Ann B. Lee and David Mumford. Occlusion models for natural images: A statistical study of scale-invariant dead leaves model. International Journal of Computer Vision, 41(1,2), 2001.

D. Marr: VISION: A computational Investigation into the Human Representation and Pro- cessing of Visual Information. W. H. Freeman and Company, New York, 1982.

B. Matern: Spatial variation : Stochastic models, their applications to some problems in forest surveys and other sampling investigations, Statens Skogsforskningsinstitut, 1960.

D. Mumford. Empirical investigations into the statistics of clutter and the mathematical models it leads to. A lecture for the review of ARO Metric Pattern Theory Collaborative, 2000.

E.P. Simoncelli, Higher-Order Statistical Models for Visual Images, Proc. IEEE Signal Processing Workshop on Higher Order Statistics, pp. 54-57, June 1999.

M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky. Random cacaeds on wavelet trees and their use in analyzing and modeling natural images. Applied and Computational Harmonic Analysis, 11:89123, 2001.

G. Winkler, Image Analysis, Random Fields, and Dynamic Monte Carlo Methods. Springer, 1995.

S. C. Zhu, X. Liu, and Y. N. Wu. Statistics matching and model pursuit by efficient MCMC. IEEE Transactions on Pattern Recognition and Machine Intelligence, 22:554569, 2000.

S. C. Zhu, Y. N. Wu, and D. Mumford. Minimax entropy principles and its application to texture modeling. Neural Computation, 9(8):16271660, November 1997.