HEURISTIC INFERENCE IN SENSORY PERCEPTION WITH AN APPLICATION TO COMPUTER VISION[1]

Ulf Grenander

SEPTEMBER 2002

In 1948, when Norbert Wiener published his path breaking work *Cybernetics: Communication and Control in the Animal and the Machine*, computer technology was in its infancy. Most computers that were needed to implement his ideas were of analog type, certainly in animals but also in machines. They were barely enough for executing simple algorithms, for example for extrapolating stationary time series or for control of servo mechanism. One can only speculate what Wiener's genius would have accomplished with todays computing machinery with their speeds in the Giga and Tera Hz ranges.

Nonetheless, the book caused a paradigm shift in several sciences and technology, in particular in biology/medicine. Not that his ideas were accepted uncritically by the scientific establishment; on the contrary they were met with skepticism and lack of appreciation. It was only with generational change that the cybernetic view was gradually accepted. Indeed, there was a period in the 1960's and 1979's when it appeared that steam had gone out of the movement, but later on it returned with renewed vigor, often with no explicit reference to the term. Today much work in the cognitive and neural sciences is dominated by Wiener's communication/control scheme, albeit in forms changed due to changes in technology. It is striking how similar are the thoughts that influence studies in animals as in machines.

Two great contemporaries of Wiener, Claude Shannon and S.O.Rice, created the fields of information theory (19..) and statistical signal processing (19..), tools that play a major role in bio/medical research, also in a cybernetic mode. Signals genmerated by sensory inputs are studied by these methods, in particular single signals from an individual cell, but recently multi-electrode techniques have made it possible to analyze several signals together. This

[1]Supported by NSF-DMS-00774276

points toward understanding of global mental functions but our present knowledge is scanty.

To *understand understanding* seems to require different concepts, more global ones, that abstract from the detailed specification of neural signals. In the absense of hard empirical knowledge about total brain function it is tempting to speculate about that which we know little or nothing about. Following Oscar Wilde's dictum, we shall fall for this temptation, and follow innumerable philosophers, novelists and poets since time immemorial, but we shall do this in a formal manner, expressing thoughts,doubts, love and hate and other feelings in mathematical language, while keeeping in mind the complexities of the human mind, its ambiguities with its illogical and inconsistent features.

Our starting point will be metric pattern theory [2] in which regular structures are constructed using generators and connectors, to build probabilistic knowledge representations of the phenomenon being studied, in the present case mental processes. We shall attempt to do this in the following PART I, where the subject is not information but knowledge = structured information, and heuristic processing of knowledge in terms of pattern theoretic prior probabilities. In PART II this will be specialized to computer vision for images acquired by range sensors. There we shall try to *imitate processes in the animal mind by processes in the machine*, to wit, automated understanding of range inmages. If we can do this we have shown, not that animal mind processes are governed by certain algorithms, but that those algorithms are consistent with what we experience subjectively when we deal with qualia. Hence we do not claim to have established any principles empirically, only to have shown their plausibility. These principles will expressed through statistical heuristics.

This paper is a continuation of Grenander (2002b).

A reader who is not upset by the speculative nature of the following may want to take a look at Grenander (2002a), even more adventurous!

PART I

General Statistical Heuristics for Perception

1 Specifying the Information Status of Perception.

Any adequate formulation of the problem requires a careful description of the prior knowledge about the world scene in question and of the means available for observing it: the information

[2] Pattern theory is presented in Grenander (1993), referred to as GPT

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:

$$\mathbf{K} = \{k_1, k_2, k_3, ...\} \tag{1}$$

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; see also Grenander(2002). 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 $<$.

1.1 Examples.

MACHINE: 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

PART II. 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

This knowlwdge status was assumed in Grenander (2002b).

ANIMAL: With human perception consider the very limited Knowledge Box

$\mathbf{K^{(4)}} =$

knowledge element	element descriptor
k_1	visual input
k_2	auditory input
k_3	intelligence: an animal is likely to be in the scene; parameters g,h,...

This concerned external perception, but we can equally well have internal activities:

$\mathbf{K^{(5)}} =$

knowledge element	element descriptor
k_1	hunanM1 present in mind
k_2	humanF1 present in mind
k_3	humanM2 present in mind
k_4	humanM1 courting humanF1
k_5	intelligence: humanM2 in love with humanF1

Such knowledge structures were used in Grenander (2002a).

2 World Model= Generators+Connectors+Priors .

The organization of algorithms for understanding will be based on a theory of the world in which the animal/machine lives. Indeed, without any theory predicated it seems impossible to organize and analyze the received perceptions in a meaningful way. We shall express the theories in pattern theoretic form, see GPT, PART I.

2.1 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.

EXAMPLES (MACHINE):

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

(ANIMAL):
$auditory < talk < answer$
$visual < color < yellow$
$feeling < anger < jealousy$
as used in Grenander (2002a).

2.2 Invariance via Similarity Groups.

Patterns are formed as equivalence classes of images w.r.t. a similarity group S; $\mathcal{P} = \mathcal{I}/S$.

MACHINE): 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

(ANIMAL): The patterns are representing mental entities expressing external inputs, thinking, feelings, external outputs...

(a) Similarity groups are of the form $S_1 \times S_2 \times ...S_m \times ...$ where S_m means the permutation group over the modality M_m. See Grenander (2002a) for details and examples. A good illustration in Figure 2.5.1, same reference.

2.3 Connectors.

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

EXAMPLES (MACHINE):

trunk side 1 $\leftrightarrow$ *trunk side* 2
vehicle $\leftrightarrow$ *ground*
vehicle body $\leftrightarrow$ *wheel*

(ANIMAL):
modality $\downarrow$ *modality* 2
as explained in Grenander (2002a) with lots of cases.

2.4 Probabilities.

Any realistic inference theory for perception, whether for animals or machines, must be probabilistic to account for the high variability in the observed images (in the sense of GPT). But the interpretation of the used probabilities can be different for animals and machines. For animals we have the added uncertainty due to limitations in predictability of their mental processes.

(MACHINE): 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 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 wonderful work of Matern (19??). 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. (ANIMAL): For animals the

situation is different. It is less clear how to find the probabilities for a certain animal in a certain situation to react neurally - how would the animal acquire the needed probabilities ? It seems unnatural to assume that they be obtained by some explicit measuring process; nature does not operate like an efficient engineer optimizing behavior by precise learning algorithms. At least we believe that this is not so.

Instead we shall think of such a probability as just another parameter that controls neural behavior and one that is not necessarily defined in the frequency sense. We shall operate heuristically with the probabilities. For example, given two functionals $f_1(I^{\mathcal{D}})$ and $f_2(I^{\mathcal{D}})$ of the observed image we shall treat them as if they were independent stochastic variables, at least f_1 is very different from f_2 in nature. This is vague and intentionally so. It reflects our view that such randomness should be thought of operationally: they are adequate if they lead to behavior that is more or less consistent with what is being observed. No doubt this view will appear controversial, to say the least, but we would like to appeal Einstein - in science one should try to express onself precisely but not more precisely than one's own thinking.

3 Statistical Heuristics.

The heuristics will be based on a subset of the knowledge that is available. The selection of the subset is assumed to be done during evolution of the species and development of the individual in animals. For machines this is done of course by the designer and will be more arbitrary.

3.1 Indicators.

The subset will consist of *indicators*, that is functionals $\omega_1, \omega_2, \ldots \omega_N$, that are assumed to compress much of the available knowledge that is useful for the intended inference. The functionals need not be scalar valued. The whole set is denoted

$$\Omega = \cup_1^N \omega_k \tag{2}$$

(MACHINE): For example

(i) ω_1= 1D histogram

(ii) ω_2= empirical spectral density

(iii) ω_3= boundary values B from the boundary of $I^{\mathcal{D}}$, the observed picture

(iv) ω_1= regions of approximately constant B-values

(v) ω_5= intelligence about scene

(ANIMAL): For example

(i) ω_1= talk heard by the observer

(ii) ω_2= gestures seen

(iii) ω_3= movement observed

(iv) ω_4= impression of facial look

(v) ω_5= body language

(vi) ω_6= intelligence about the emotional situation

......

3.2 Inference based on indicators.

Inference shall be of the form

$$\phi : \Omega \to \mathcal{P} \tag{3}$$

cence

Some possible types of inference:

(i) prediction of the next image to be expected, perhaps several possibilities with attached probabilities

(ii) extrapolation of an interior part of the image that was not observed (fully); this is of importance for reasons discussed in Grenander (2002b).

(iii) analysis, including the estimation of parameters in the regular structures that make up the pattern class P so that it takes the values of the indicators into a pattern. The mapping ϕ can be deterministic or random.

4 An Example of Statistical Heuristics in Animal

Prediction can take the form $image_1 \rightarrow image_k, probabolity = p_k$. For example

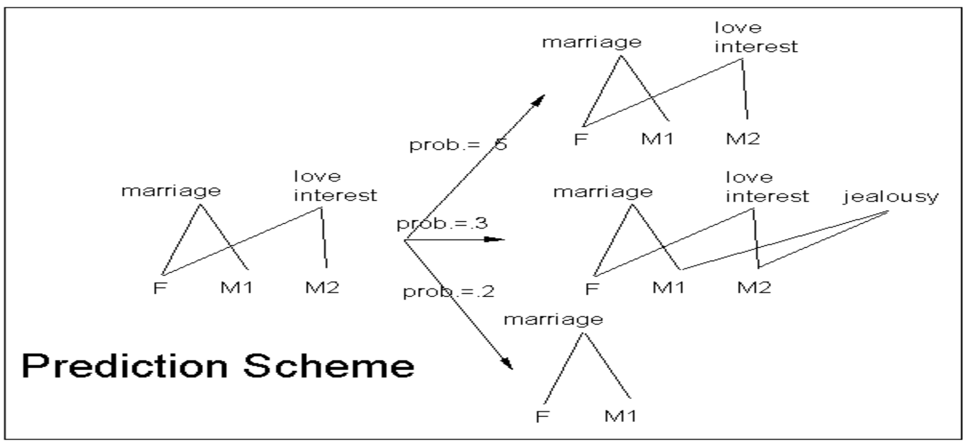

Figure 1

where an eternal triangle leads to three possible reolutions with specified probabilities. More about this in a future report.

PART II

Special Statistical Heuristics for Range Vision

5 Pattern Theoretic Approach to Forest Scenes.

We shall concentrate on interpolation inference. 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). Put a frame of widt $(L1, L2)$ 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 $F = I^{\mathcal{D}}_{ext} - I^{\mathcal{D}}$ 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? To answer such a question we must know something about the background and that is exactly the task of the interpolator.

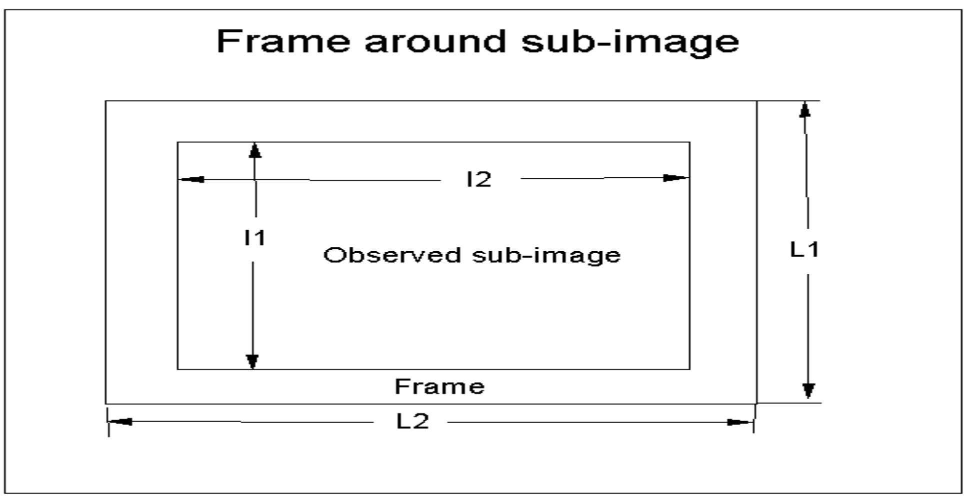

Figure 2

In Grenander (2002b) we attempted various interpolation schemes, in particular those based on minimizing the conditional energy

$$E_{cond} = \sum_{i_1=1}^{l_1} \sum_{i_2=1}^{l_2} e[I^{\mathcal{D}}(i_1, i_2)] \tag{4}$$

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

$$e[I^{\mathcal{D}}(i_1, i_2)] = f[I^{\mathcal{D}}(i_1+1, i_2)] + f[I^{\mathcal{D}}(i_1-1, i_2)] + f[I^{\mathcal{D}}(i_1, i_2+1)] + f[I^{\mathcal{D}}(i_1, i_2-1)] - 4f[I^{\mathcal{D}}(i_1, i_2)] \tag{5}$$

We used the Bessel K densities b and

$$f(x) = log[b(x, c, p)] \tag{6}$$

Here we have used

$$b(x, c, p) = \frac{1}{\sqrt{\pi}\Gamma(p)} (\frac{c}{2})^{\frac{3p}{2}-\frac{1}{4}} x^{p-\frac{1}{2}} K_{p-\frac{1}{2}}(x\sqrt{\frac{2}{c}}) \ .$$

For the Gaussian case $p = +\infty$ we get the classical harmonic function interpolation. 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 $p = 1$. The results were disappointing. The reason for this poor inference performance is of course that the assumed pattern probabilities do not catch much of the real image structure. Indeed, it says only that all the differences

$$I^{\mathcal{D}}(i_1+1, i_2) - I^{\mathcal{D}}(i_1, i_2), I^{\mathcal{D}}(i_1, i_2+1) - I^{\mathcal{D}}(i_1, i_2) \tag{7}$$

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 are described . Of course we have also derived 3D approximations, but while the 2D Bessel K approximations provided highly accurate quantitative agreement with dat. this was not the case for 3D; only qualitative similarities with data were observed.

To get better results we must include more knowledge about 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.

6 Heuristics for Forest Scenes

We shall use the generators from section 2.1: $G = \{foliage, ground, trunk, sky\}$. This generator space is a bare minimum and may have to be increased but 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 3 Note how stretches of nearly constant range values or linearly increasing ones are separated by rapidly changing values.

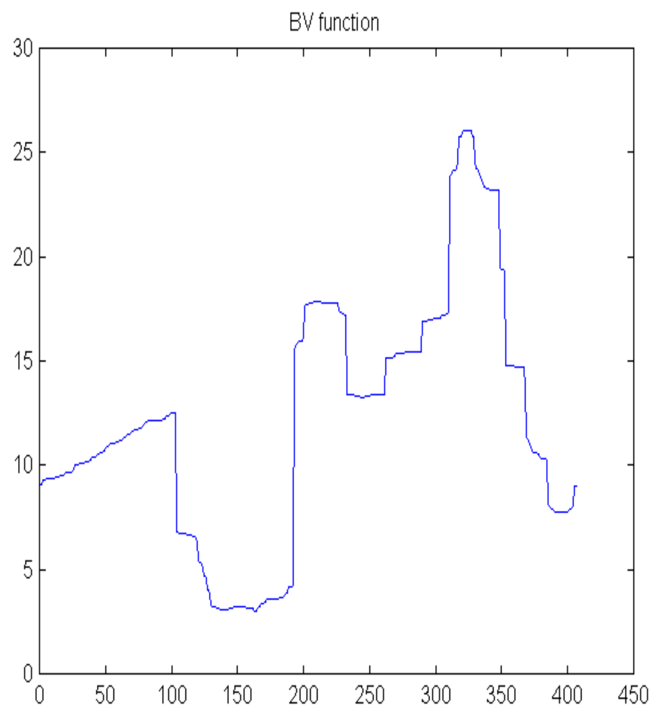

Figure 3
Determine

$$M = max_s BV(s); m = min_s BV(s) \tag{8}$$

and introduce the range levels

$$[r_k, r_{k+1}]; k = 1, 2, ...N; r_k = m + (k-1)/N(M-m) \tag{9}$$

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 $\{s : r_k \leq BV(s) < r_{k+1}\}$. Now reject small intervals with lengths less than some thresold value and filling in holes of length smaller than some other threshold. This gives us a set INTERVALS = $\{int_1, int_2, int_3, ...n_{int}\}$ of these intervals, each interval $int \in INTERVALS$ associated with some range levels and written as $[p_1(\nu), p_2(\nu)]$. We shall use modular addition $n_{int} + 1 \equiv 1$.

We shall need the following concept. Define a function

$$Q(s) = 1, s \in BELOW; Q(s) = 2, s \in RIGHT; Q(s) = 3, \in UP; Q(s) = 4, s \in LEFT \quad (10)$$

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

$$\omega_\nu^1 = LS(\nu) = line\ segment\ p_1(\nu) \rightarrow p_2(\nu) \quad (11)$$

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 $C(\nu)$ including 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 4

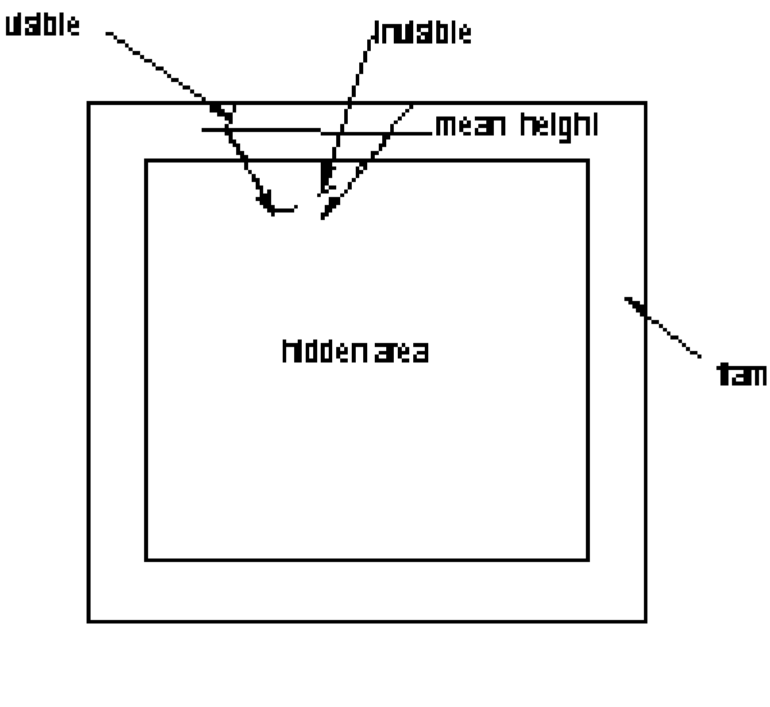

Figure 4

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 thhis question unanswered. Then we get indicators of the second class

$$\omega_\nu^2 = (U_1, U_2) \tag{12}$$

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 ither wird, we are estimating derivatives, gradients, which is a notoriously sensitive task for complex pictures with much local variability. This is the purpose of the vectors U_1, U_2.

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

$$I^{\mathcal{D}}(i_1, i_2) = 0 \; redefined \; as \; I^{\mathcal{D}}(i_1, i_2) = M \tag{13}$$

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

6.1 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

$$r_{parallel} = R(\|U_1 - U_2\| \leq \epsilon_1) \tag{14}$$

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

$$r_{vertical} = R(\|(U_1 + U_2)/2 - col(1,0)\} \leq epsilon_2) \tag{15}$$

. The value r_1 is 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

$$\phi_1(I^{\mathcal{D}}) \rightarrow \text{"}trunk \; element \; at \; (p_1(\nu) + p_2(\nu))/2\text{"} \tag{16}$$

. 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 $int(\nu)$. Implicitly this makes an indepence assumption following our heurisic principles. With the complementary probability we do nothing.

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

6.2 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^*, again 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 will not be includede 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 5, forming a Markov process of order 2. The process continues as long a the line segments have positive length. We access to the observables p_1, p_2, U_1, U_2.

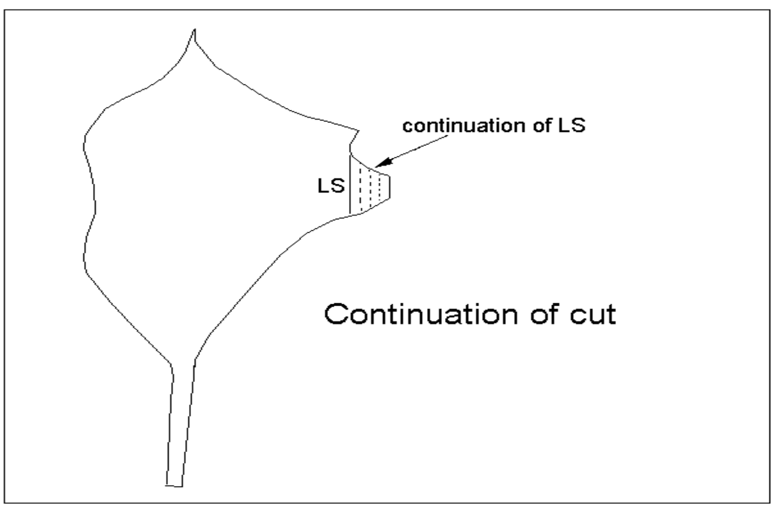

Figure 5

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, using discrete space, that the cuts along the following cuts forms a Markov process satisfying the coupled Langevin SDE's

$$x_2^1(t) = 2x_2^1(t-1) + x_2^1(t-2) - k * [x_2^1(t-1) - x_2^2(t-1)] + e_1(t) \tag{17}$$

$$x_2^2(t) = 2x_2^2(t-1) + x_2^2(t-2) + k * [x_2^1(t-1) - x_2^2(t-1)] + e_1(t) \tag{18}$$

with $e_1(t), e_2(t)$ as 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 $x_2^1(0) = p_1, x_2^2(0) = p_2, p_2 < p_1; x_2^1(1) = q_1, x_2^2(1) = q_2$. 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 6 we show some set continuations using the *Markov cut model*; primary cut in red. Note how the direction tendencies at the initial cut propagate into the protuberance but gradually die out. The speed of this depends upon k and regulates the (random) size of the protuberance, while σ^2 controls the smoothness of the boundary.

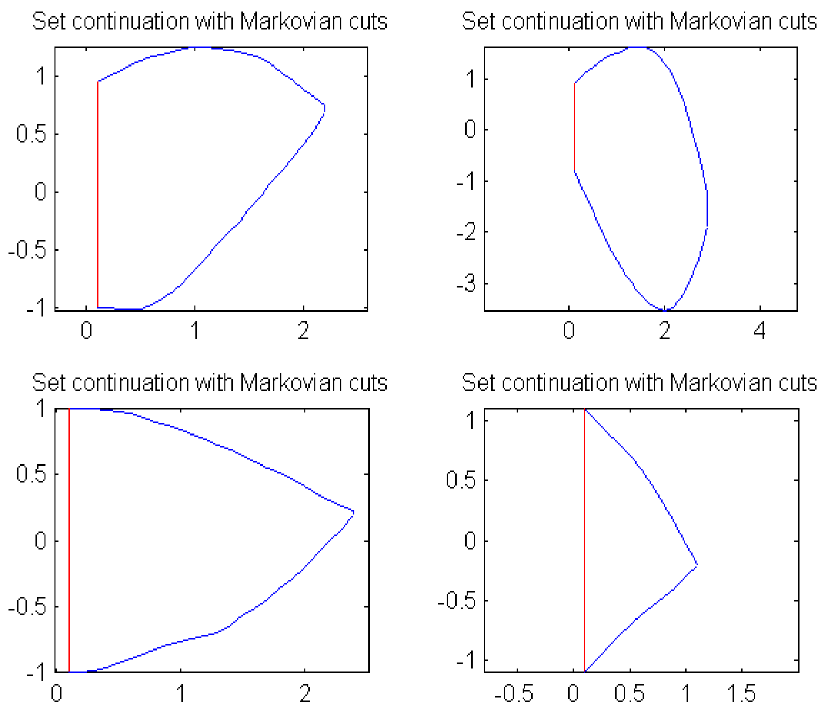

Figure 6

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.

6.3 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 probability

$$r_{same\ height} = R(|P_2 - P_3\| \leq epsilon_3 \; and \; \|P_4 - P_1\| \leq \epsilon_3) \tag{19}$$

Where R is some 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.

6.4 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

$$I^* = min_{t \in T} I_t^* \tag{20}$$

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

7 Software for Heuristics Applied to Range Images

The algorithm that we have described is fairly complicated with many steps and conditions. Therefore we have coded it in a modular way with a main calling several auxiliary programs. In spite of its length it runs at about the same time as the harmonic function interpolator that we ruled out as useless.

7.1 Evolution of the Heuristics.

While writing the software the many modules of the software developed in time from a primitive beginning that did not work well at all. To improve the performance, code segments were added or modified, leading to successively better behavior. Hence, we did not proceed in the ideal engineering way, starting from an optimality criterion and deriving the optimal algorithm from it. Our procedure was admittedly less elegant, more heuristic, but perhaps also closer to how perception tools have evolved in nature over time, correcting poor properties and aiming for, if not perfection, at least acceptalble performance.

The present stage of the algorithm should be considered as only one step in the evolution evolution. It realizes the ideas in PART I but specializes some of the choices and of course the parameters. The parameters are also thought of as evolving in time. The algorihtm has been executed hundreds of times, successively improving the inferences.

7.2 Perceptions.

Let us look at some of the results.

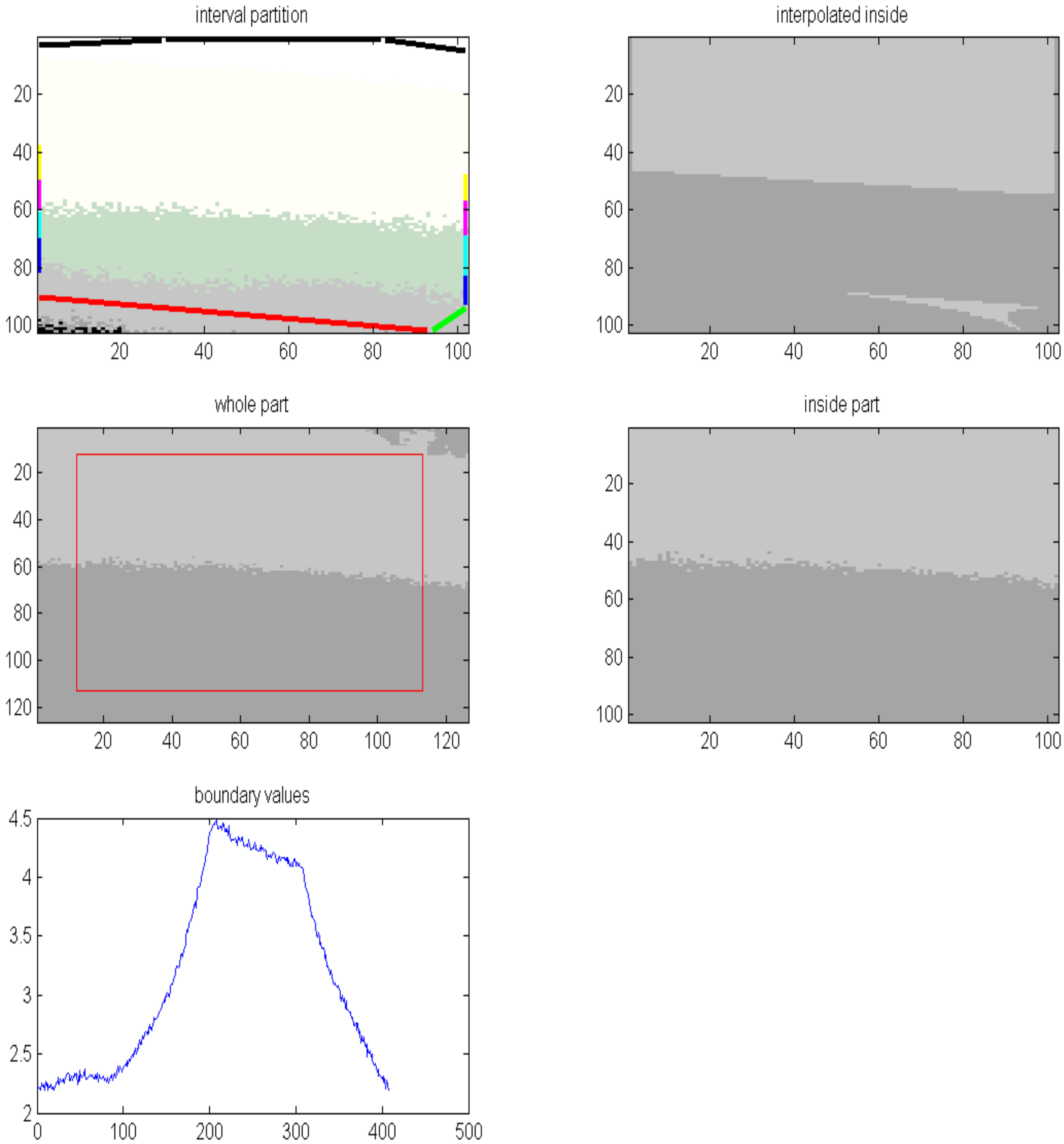

Figure 7

Figure 7 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 one 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. The upper right panels is the inference, here 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.

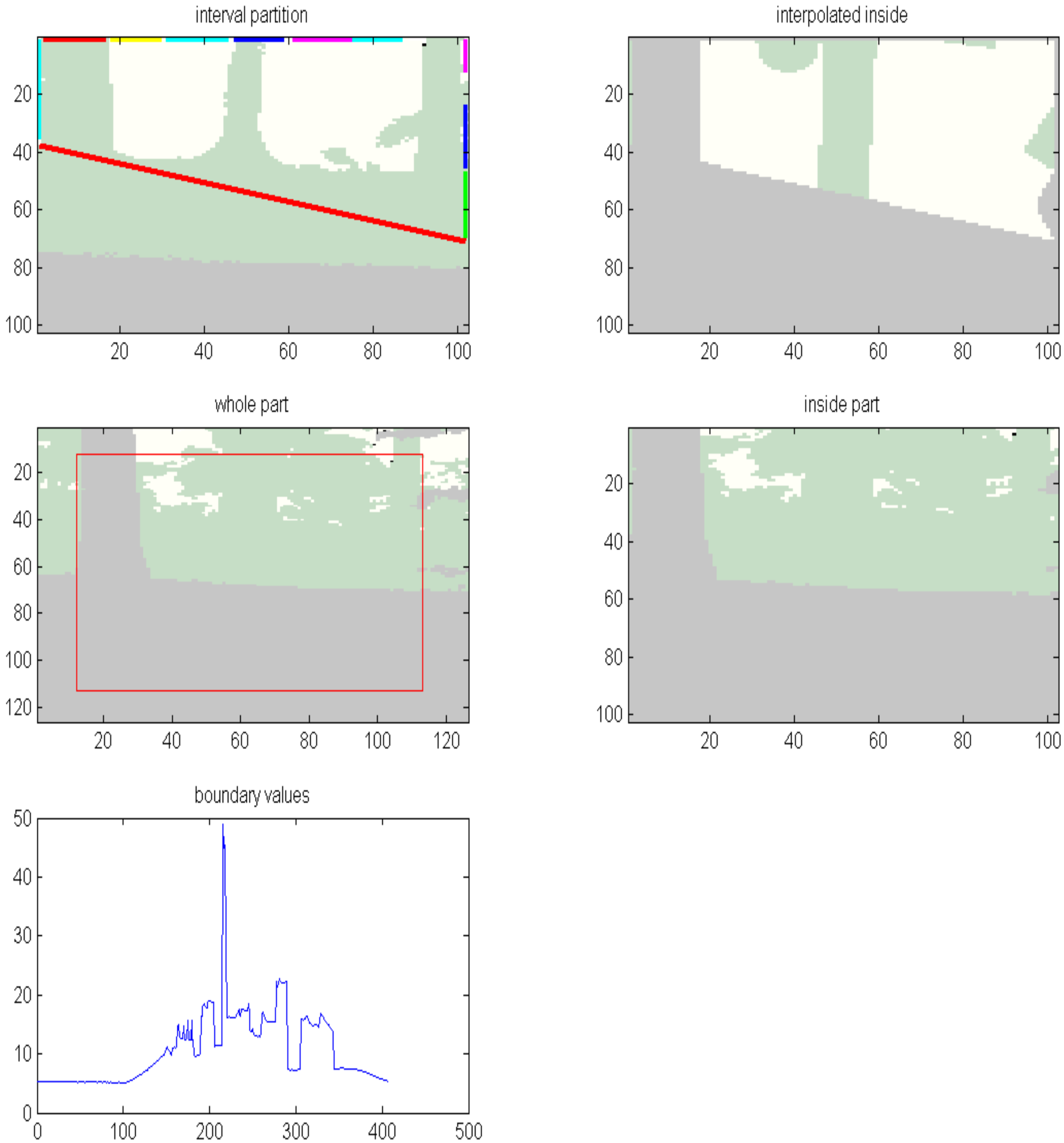

Figure 8

This scene has a big tree and another further away. The inference is fairly good.

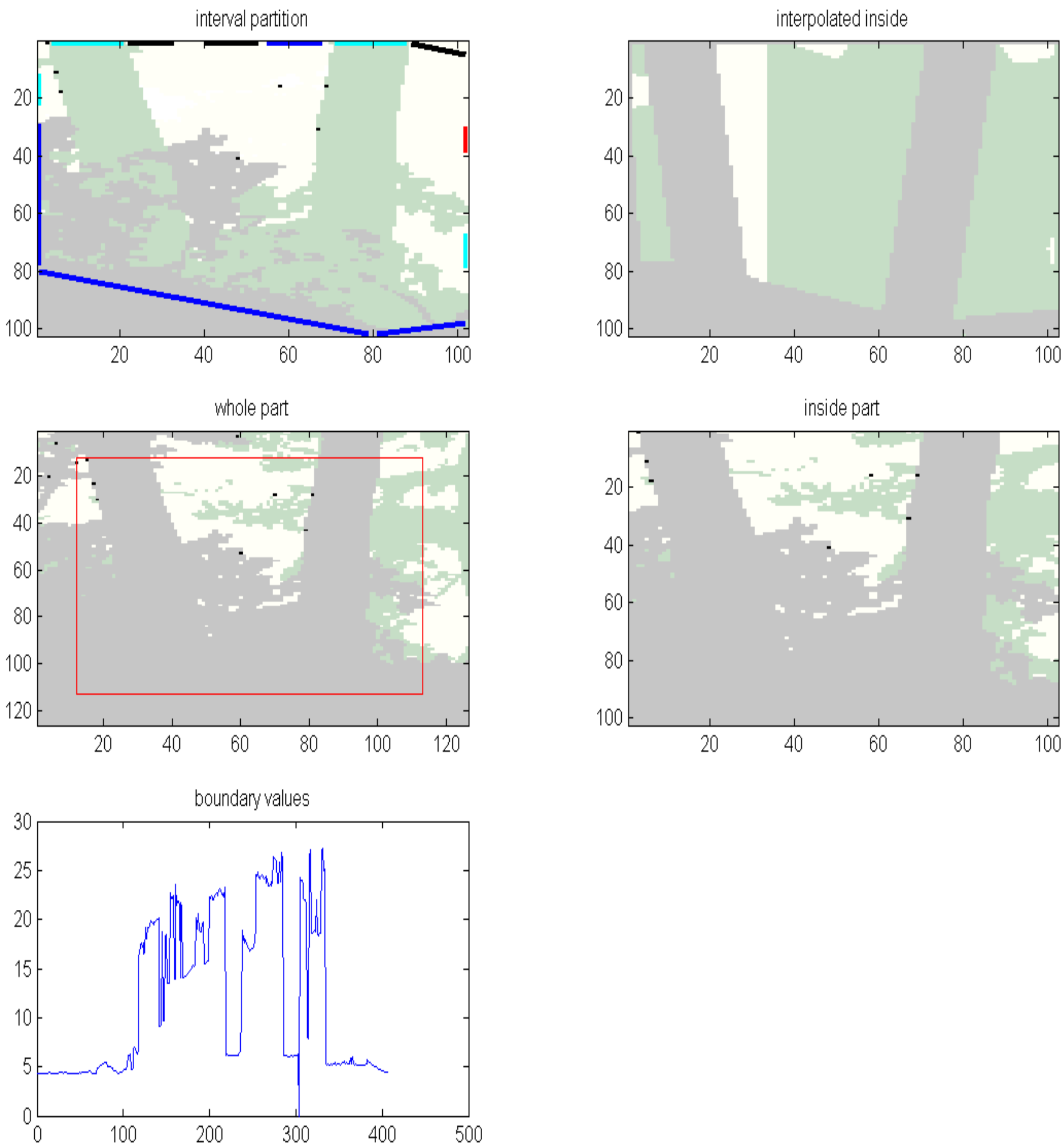

Figure 9

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.

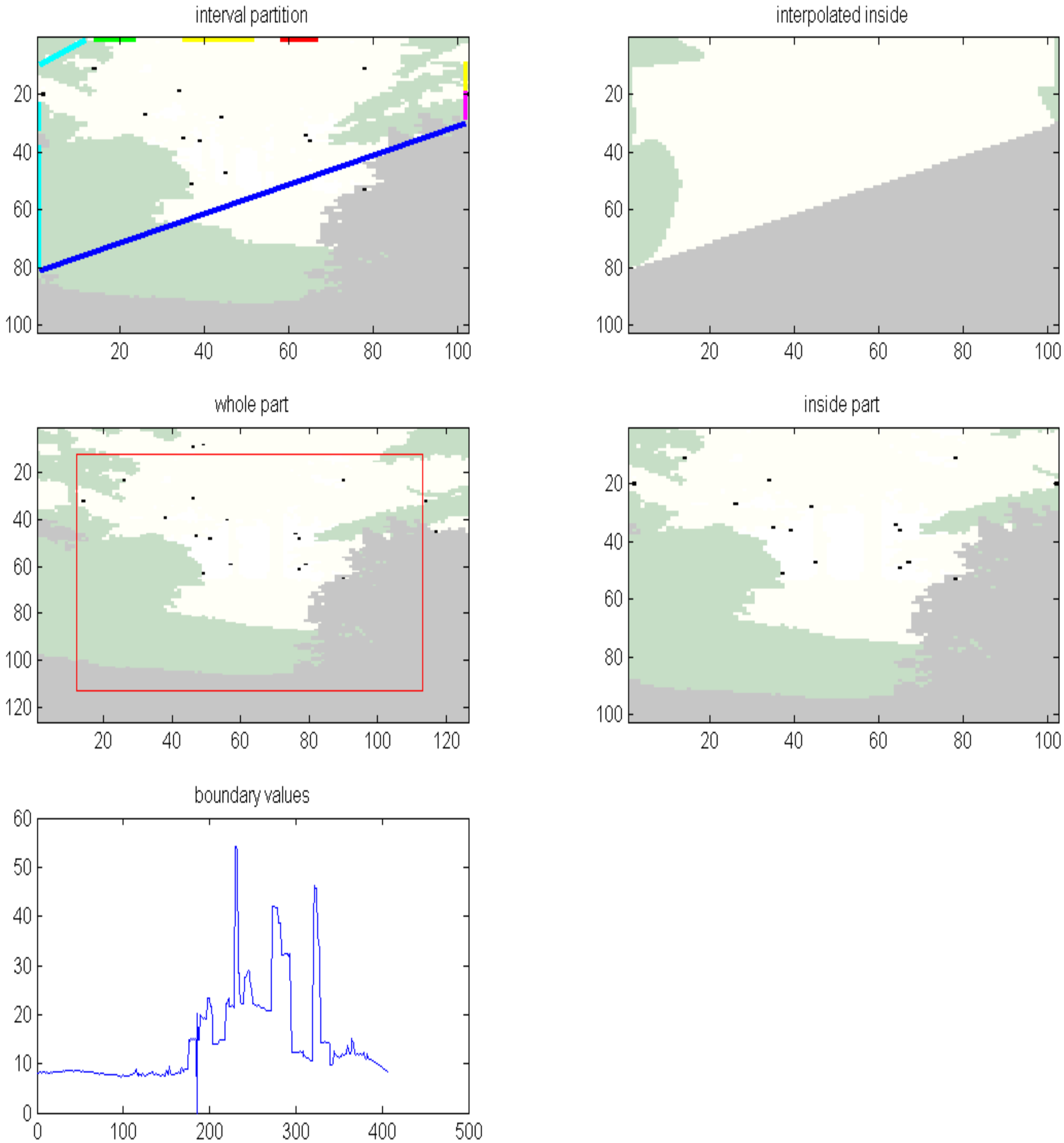

Figure 10

Here big foliage 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.

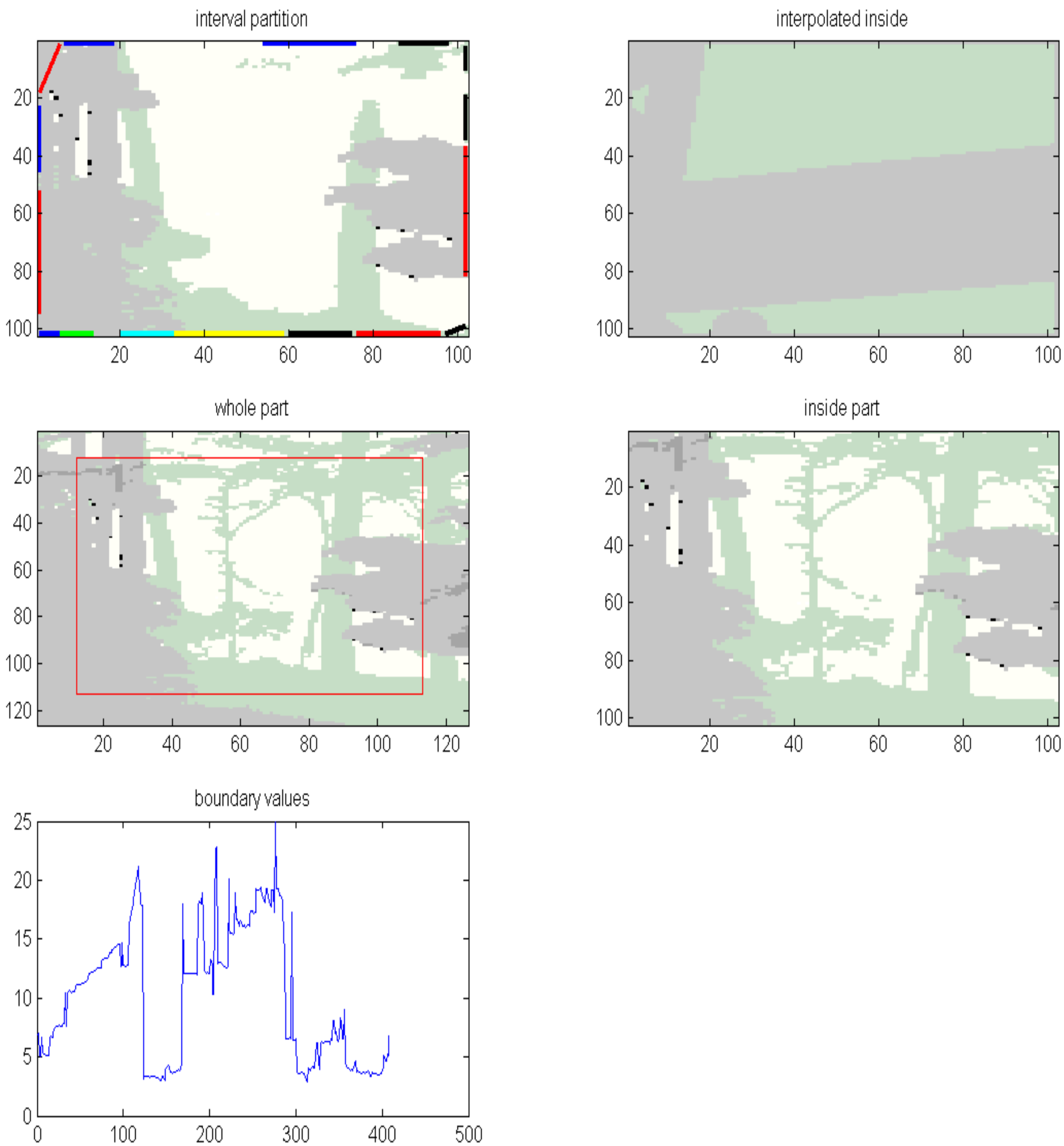

Figure 11

A similar effect: the right foliage has been continued much too far, the direction tendencies have been poorly estimated.

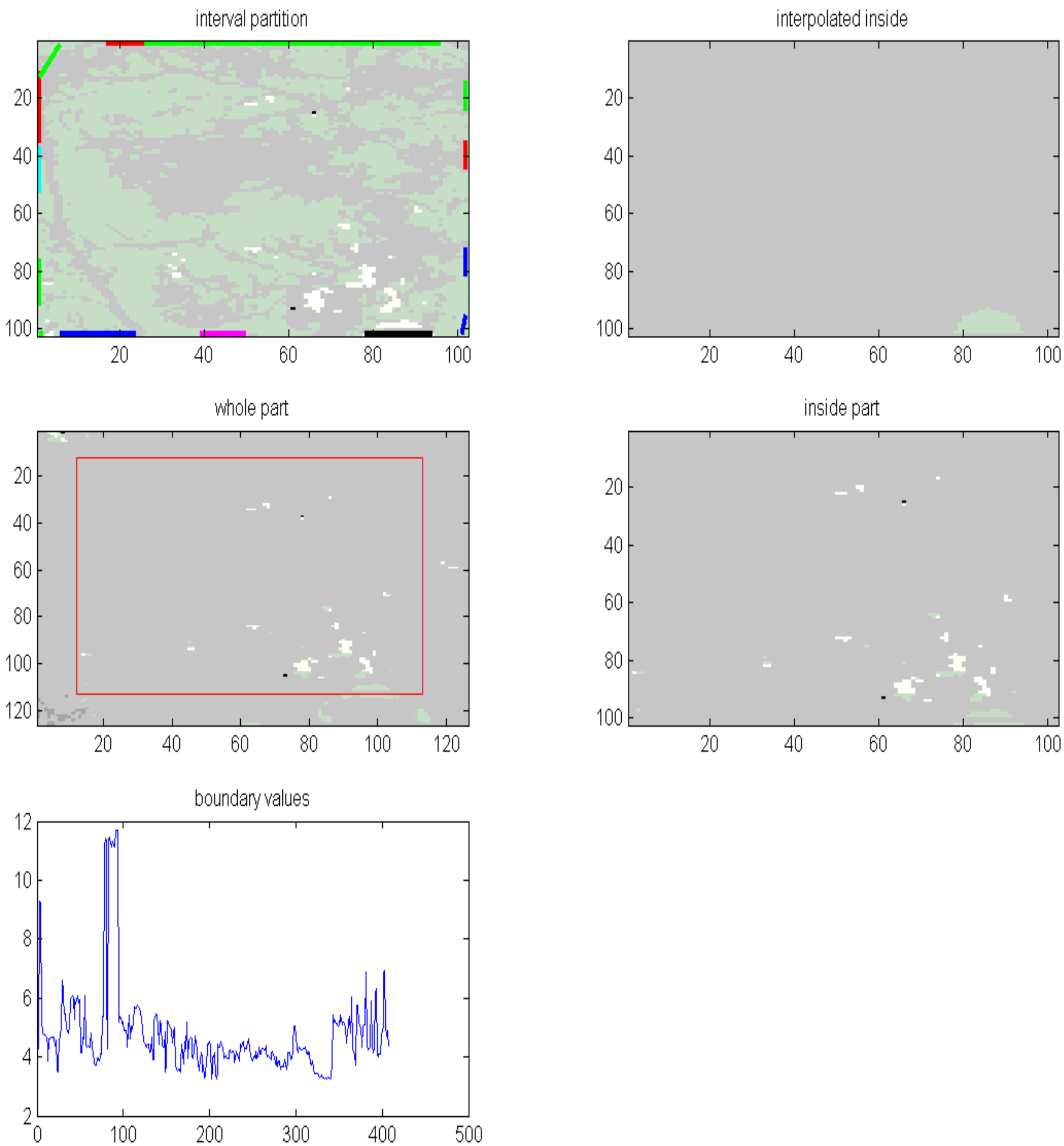

Figure 12

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

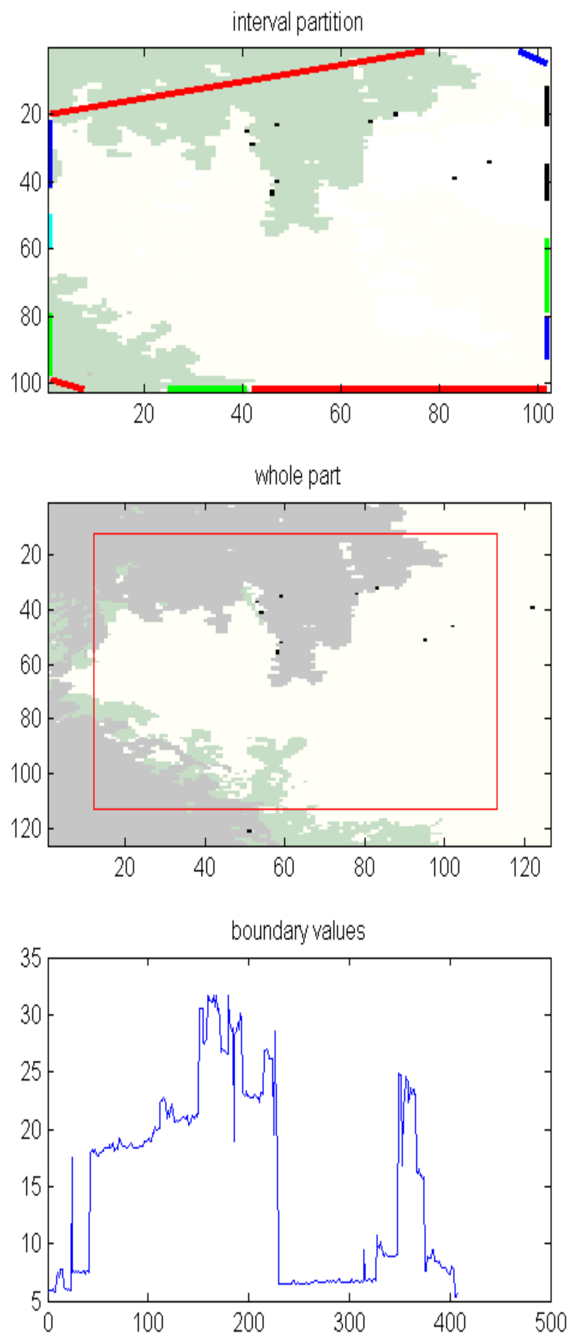

Figure 13

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

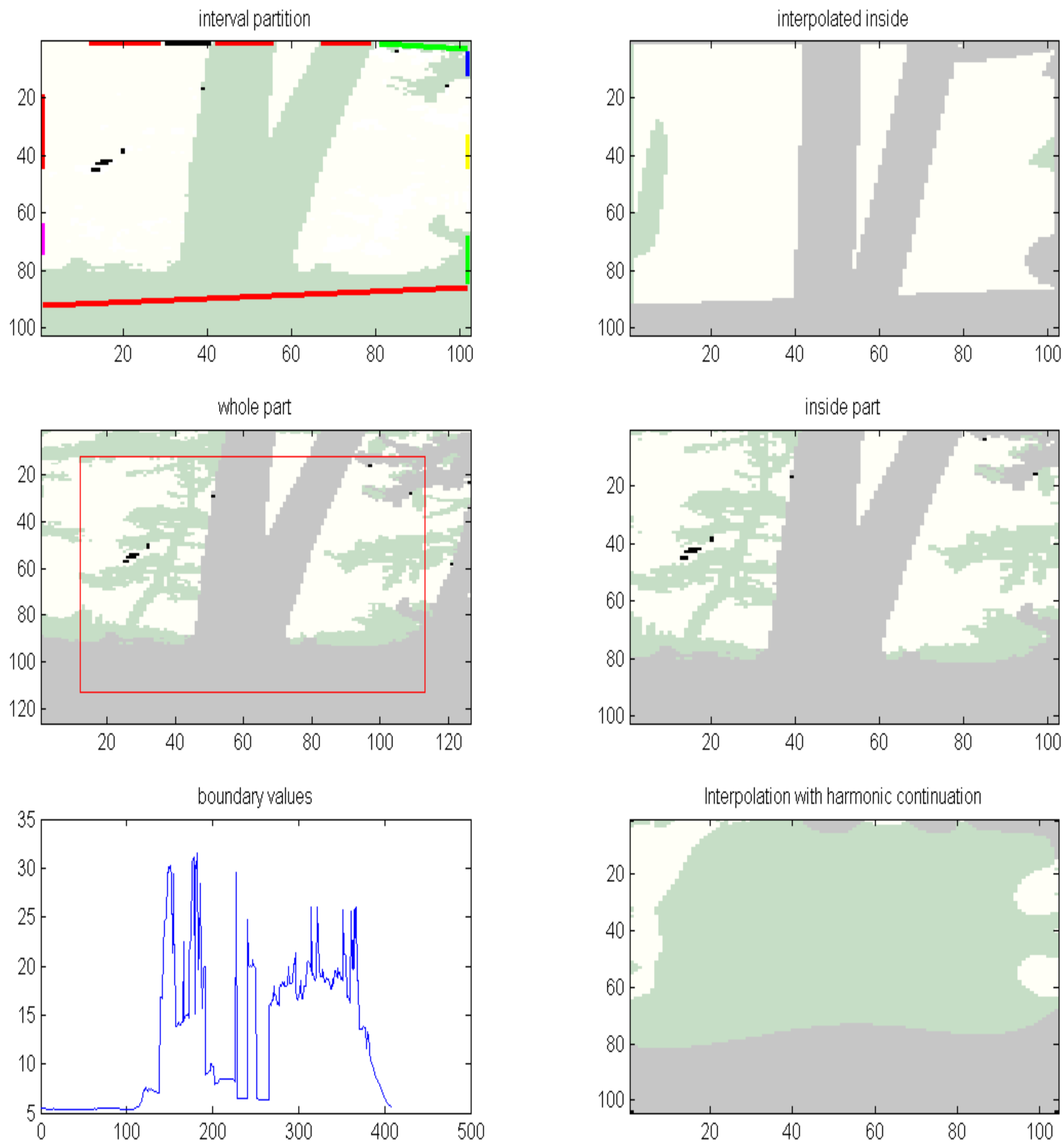

Figure 14

Almost perfect perception inference. For comparison the lower right panel shows the harmonic function interpolator. As we already know its knowledge about this sort of scenes is almost nil.

7.3 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. No doubt it could be handled better by improving the modules of the code for foliage recognition, but it can be questioned if any algorithm could perform really well for foliage. Could one derive lower bounds for the error probabilities?

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 ones based on the boundary alone; the exploitation of trunk-foliage-sky-ground structure has helped a lot! At least this approach has shown its strength and deserves further study and improvement.

8 Postscript

The above study has been dominated by the dichotomy MACHINE VS. ANIMAL. Gradually, however, this contradistinction seems to vanish: the ideas are very similar for both instances. Indeed, we have used the same pattern theoretic paradigm to represent both automated machine understanding and to human mental processes. This point to a more complete coalescence of ideas in the future.

Once we have deepened this approach, both conceptually/analytically and computationally, it should be applied more systematically to the study of human thought inspired by what we have learnt about machine understanding. Hence, we announce a more finished version of the program on the CD "Windows on the World":

(i) To extend the ideas in Grenander (2002a) along the lines indicated above and with a generator space of more realistic size, perhaps 10000 - 100000 with corresponding increase in modality structure.

(ii) To code the mind structure in a more systematic and efficient ways, perhaps in C++.

(iii) Include the new memory representation in the code.

(iv) Apply and modify the codes according to different psychological doctrines.

(v) Study the interaction of two different mind structures in the same world.

(vi) Consider the possibility of implementing the codes by a neural network.

At present this will seem over ambitious to some, impossible to many, but we believe that such a venture will succeed. It can be done - it will be done.

REFERENCES

U.Grenander (2002a): Patterns of Thought, www.dam.brown.edu

U.Grenander (2002b): Computer Understanding of Natural Scenes, www.dam.brown.edu

O. Husby and U. Grenander (2001): A Model for Recognition of 3D Non Dense Objects in Range Images, www.dam.brown.edu

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

S. O. Rice (1944), "Mathematical analysis of random noise," Bell Syst. Tech. J., vol. 23

C. Shannon (1948): A Mathematical Theory of Communication in the Bell System Tech. Journal

N.Wiener (1948): Cybernetics - Control and Communication in the Animal and the Machine, MIT Press