MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

A Comparison between Top-Down and Bottom-Up Image Analysis in Terms of the Complexity of Searching a Problem Space

Ali A. Abkar *
Geoinformatics Department
International Institute for Aerospace Survey and Earth Science (ITC)

Nanno J. Mulder †
Department of Electrical Engineering
University of Twente

Abstract

In the context of the analysis of remotely sensed data the question arises of how to analyze large volumes of data. In the specific case of agricultural fields in flat areas these fields can often be modeled in terms of geometric primitives such as triangles and rectangles. In this case the options are classical i.e. bottom-up, starting at the pixel level and resulting in a segmented, labeled image or top-down, starting with a model for image partitioning and resulting in a minimum cost estimation of shape hypotheses with corresponding parameters.
We report on an investigation of the search effort needed for resolving a simplified segmentation problem of partitioning an image into two segments. Experimental factors are edge length and overlap of monospectral probability distributions of two classes.
The method for quantifying the complexity of an approach is to determine the number of possible solutions at each stage in the process and the convergence rate towards a final solution of the segmentation and labeling problem.

1 Introduction

In image analysis of remotely sensed data the usual approach is bottom-up in the sense that the analysis starts at the pixel level and the end result is a segmented and labeled set of image samples. Increasingly, this is followed by conversion to vector (polygon) format based on the assumption that the edges between objects are piece-wise linear.
We have argued [1, 2, 3] that in applications of Remote Sensing (RS) to man-made objects such as agricultural fields there is sufficient knowledge about the shape of the image segments to allow top-down image analysis. Our pure top-down analysis starts with a set of shape hypotheses such as the object is-a {triangle, rectangle, circle-segment} or a combination of these according to the application (set operators on these primitives). The role of the RS data is to evaluate the current set of hypotheses and shape parameters and to modify this set until a minimum cost instantiation is found.
Counter arguments to our "pure" top-down approach have centered on the assumed complexity of the shape of agricultural fields and on the large overlap amongst the clusters of multispectral data between clusters of different classes.
In this paper we argue that if approximation of 2-dimensional object shapes by polygons is accepted then any object shape in the same context can be approximated by a union of triangles. Adjacent triangles of different objects share a straight "edge"; therefore, locally the problem of hypothesis generation and testing is reduced to the hypothesis of samples of class A and class B in a region of interest with a linear decision function for image sample membership. The linear membership decision function has two parameters for position and orientation (par1, par2) plus an edge length which defines the size of the region of interest.
We compare the top-down and bottom-up approaches for the simple case of a linear edge with experimental variables: overlap of probability density functions and edge length. Complexity is defined in terms of the number of possible solutions in a problem space and the number of operations required to reduce the number of solutions to the minimum cost solution(s).
The process of image analysis is mapped onto a problem of navigation through a split/merge tree.

1.1 Definitions

- Monospectral image: Rad(xi,yi), xi:=$0..2^N-1$, yi:=$0..2^N-1$ is a set of Dirac samples of photon counts at image sample position (xi,yi) scaled to 0..255 (N = 1, 2, 3...).
- Membership decision function:
FuncDecideClass(par1,xi,yi) = xi-par1, for which

* Address: PO Box 6, 7500 AA Enschede, The Netherlands.
E-mail: `abkar@itc.nl` or `ali@mi.el.utwente.nl`
† Address: PO Box 217, 7500 AE Enschede, The Netherlands.
E-mail: `n.j.mulder@mi.el.utwente.nl`

the parameter to be estimated is the par1 (edge position parameter) for known orientation; with

IF FDC < 0 THEN Class = A
ELSE Class = B.

- A hypothesis map Hyp(Class,xi,yi) is a map of Dirac samples of the membership decision function on the domain class = {A,B} (Boolean).
- An evidence map Evi(Class,xi,yi) is a mapping from Rad(xi,yi) to the likelihood P(Class|Rad(xi,yi)) under the assumption of equal priors P(Class) in the region of interest.
- Cost(Hyp(k,xi,yi), Evi(l,xi,yi)) = Hyp(k=A,xi,yi) × Evi(l=B|Rad(xi,yi)) + Hyp(k=B,xi,yi) × Evi(l=A|rad(xi,yi)), this the per sample expected cost of mislabeling (sum of the off-diagonal elements of the confusion matrix by the use of a unit cost matrix of Dfl. 1.00 for every completely misclassified image samples).
- Cost(Hyp(k), Evi(l)) = sum-over(xi,yi) of Cost(Hyp(k,xi,yi), Evi(l,xi,yi)), over the region of interest and eventually over the complete image.

The essence of our method is that we minimize the cost of mislabeling per object. This contrasts with the standard method of minimum error or maximum likelihood per sample (pixel).

2 The complexity of top-down edge detection

Complexity in terms of the number of possible solutions depends here on the resolution of the 2-dimensional parameter space which is related to the resolution of the image and the corresponding hypothesis map Hyp(Class,xi,yi). The minimal change in the hypothesis map with a change of parameter must be one sample but, in practice the change in the hypothesis map must be of the order 2^N-1 indicating an "area" of the solution space: minimal Cost(par1, par2) of the order 2^N+1.

Figure 1 shows a hypothesis map, an image and the graph of the cost as a function of the distance parameter (par1), which has a minimum at par1=32 and par2=0. The search history is overlaid on the parameter space (problem space) and lines of constant cost are drawn for each evaluation of the cost function. Each independent evaluation reduces the dimension of the problem space by one.

For a given dimension of parameter space (1 in this case), a very large region of interest, a known decision function and a known P(Rad | Class), it is possible to find the optimal hypothesis map (as a function of optimal parameters) in dim+1 steps [4].

In our case the top-down complexity is defined by the dim+1=1+1=2 operations needed for the optimal bisection of a region of interest. Each evaluation of Cost(Hyp(k), Evi(l)) requires 2^{2N} conditional additions (over the complete image) if no use is made of the limited area (partial derivative) comprising the subset of map Hyp(Class,xi,yi) where the hypothesis changed.

To test the complexity of the top-down analysis in the case of small edge lengths we have performed the following experiments: Vary the expected minimum cost of confusion due to overlaps from Dfl. 0.0 to Dfl. 0.5 per sample in steps of 0.05. Vary the edge length 2^N, N:= 1..8 step 1. Figure 2 shows an example of parameter estimation where N=6 and expected minimum cost is equal Dfl. 0.45 per image sample (pixel).

3 The complexity of bottom-up image analysis

In this experiment we assume bottom-up meaning, per pixel maximum likelihood classification followed by merging of 4-adjacent evidence map elements until two map-segments remain. The generality of the approach is not severely compromised by limiting the experiments to a single vertical edge. The parameter to be estimated is the shift-x parameter.

The theoretical complexity is considerably higher at the start than with the top-down method. Initially, there are $2^\wedge(2^{2N})$ possible solution maps with 2 classes compared to max. 2^{2N} solution parameters (e.g. in case of N=6, the number of possible maps is equal to $2^\wedge(2^{2N}) = 2^\wedge(2^{2\times 4}) = 2^{4096}$). The goal state has two segments.

In practice, the feasibility of the bottom-up approach depends crucially on the effects of merge operations. The factor we expected to be of the most importance is the reduction of the number of segments by merging based on 4-adjacency and maximum common boundary length weighted by the number of pixels in each segment [5].

The computational complexity of merging is of the order of the number of neighbours in the region adjacency graph plus an overhead for editing the region adjacency graph. Initially, there are $4\times(2^{2N})$ neighbours to be considered.

When the expected minimum-cost=0/sample (no overlap of probability distributions of two classes) then the result will be two segments from which the decision function's parameters can be estimated. When the expected minimum-cost=0.5/sample (complete overlap of the probability distributions) then the merging will progress at random with slow convergence and a meaningless result.

We experimented with a single edge: Vary the expected minimum cost of confusion due to overlaps from 0.0 to 0.5 per sample in steps of 0.05. Vary the edge length 2^N, N:= 2..8 step 1.

Table 1 reports the average reduction rate of the number of segments for consecutive iterations of the merge algorithms.

Figure 3 shows initial segmentation for a case of expected minimum cost is equal Dfl. 0.45/sample and segmentation into 2 segments, reached after 136 iterations of the merge operation. Figure 4 shows the result of the application of iterative region-merging to the output of the pixel based maximum likelihood classification.

4 Conclusions

Under the conditions of the experiment the top-down method always has a lower theoretical complexity than the bottom-up method as the complexity of searching a parameter space for a specific model is always less than that of searching one solution map in the set of all possible maps.
The computational complexity (number of operations to solution) varies strongly with the expected minimum cost(parameters). The more the probability densities of radiometry given class overlap, the more obvious the advantage of the top-down method, as achieving of the optimal result hardly depends on the overlap of the probability density functions. In the worst case an exhaustive search of parameter space is required.
The generation and evaluation of an evidence map and some three hypothesis maps need not require more computer time than maximum likelihood per pixel classification, followed by one pass segmentation.

References

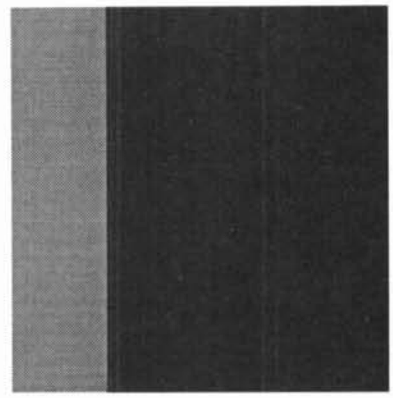
[1] N.J. Mulder, "A theory of knowledge based image analysis with applications to SAR data of agriculture". Proceedings of European Optical Society and International Society of Optical Engineering Symposium, 26-30 Sept., 1994, Rome, Italy.

[2] N.J. Mulder and Fang Luo, "Knowledge based image analysis of agricultural fields in remotely sensed images" Pattern Recognition in Practice, 1994, Vlieland, The Netherlands.

[3] Ali A. Abkar and N.J. Mulder, "Knowledge based recognition of agricultural fields in remotely sensed data" ASCI Imaging Workshop 1995, Venray, The Netherlands.

[4] N.J. Mulder, "Knowledge representation and model based image analysis of agricultural fields" ASCI Imaging Workshop 1995, Venray, The Netherlands.

[5] Ali A. Abkar, "Knowledge-Based Classification Method for Crop Inventory Using High Resolution Satellite Data." MSc Thesis, 1994, ITC, Enschede, The Netherlands.



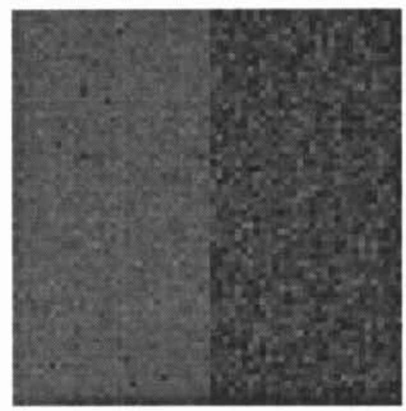
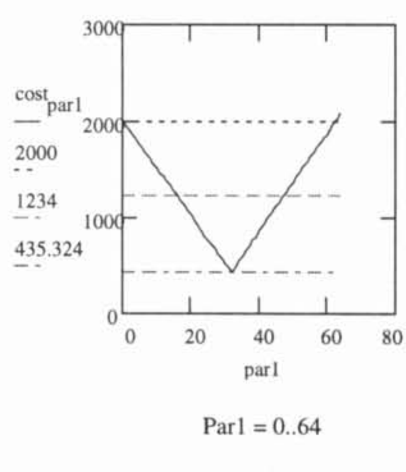


Figure 1(a) shows:
- a hypothesis map for par1=16 (shift-x) and par2=0; Grey = Class A, Black = Class B.
- the image of size 2^{2N} samples, where N=6 (64×64 samples), and the expected minimum cost is equal Dfl. 0.10 per image sample (pixel).

cost =

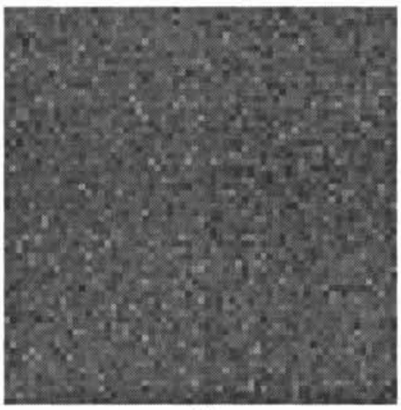
	0
15	$1.279 \cdot 10^3$
16	$1.234 \cdot 10^3$
17	$1.18 \cdot 10^3$
18	$1.13 \cdot 10^3$
19	$1.081 \cdot 10^3$
20	$1.032 \cdot 10^3$
21	979.371
22	926.551
23	874.745
24	824.533
25	774.991
26	725.834
27	679.884
28	634.076
29	583.886
30	533.468
31	486.646
32	435.324
33	491.352
34	543.806



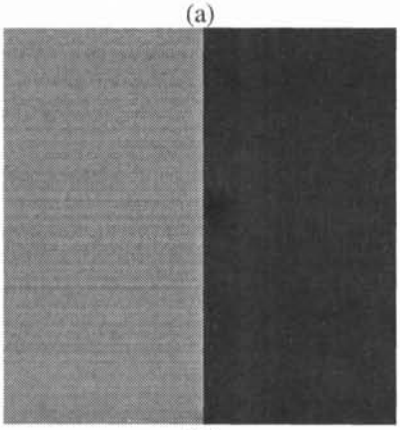
Par1 = 0..64

Figure 1(b):
- the graph of the cost as a function of par1, which has a minimum at par1=32 with the corresponding

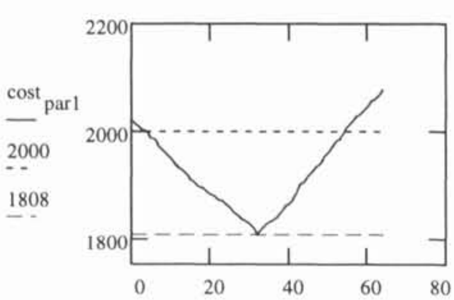
cost=435.324. The corresponding cost for the hypothesis map in (a) is 1234. The number of calculations of Cost (parameter=shift edge) is 2^N+1 (in this case 65) indicating the "area" of the solution space as shown in the graph.



(a)



(b)



(c)

Figure 2 shows: an example of parameter estimation where N=6 (64×64 samples) and expected minimum cost is equal Dfl. 0.45 per sample.
(a) the image,
(b) the hypothesis map for par1=32,
(c) the graph of the Cost(par1=shift edge) which has a minimum at par1=32.

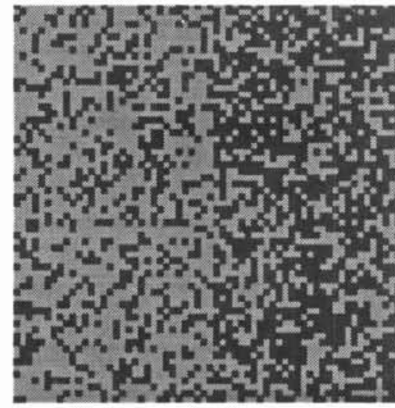


Figure 3: initial segmentation for a case of expected minimum cost is equal Dfl. 0.45 per sample with 552 initial segments. Segmentation into 2 segments, reached after 136 iterations of the merge operation.

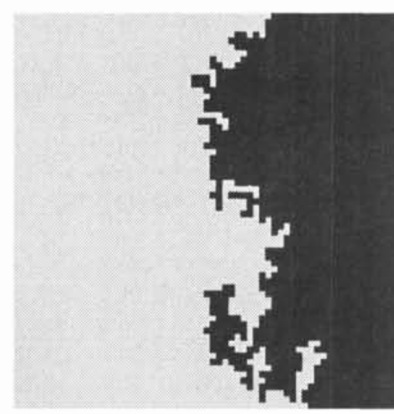


Figure 4: the result of applying iterative region-merging to the output of the pixel based maximum likelihood classification.

Iteration no	No of Segments	No of possible solutions (merges)	No of merged segments
0	4096	2^{4096}	-
1	552	551	301
2	222	221	71
3	134	133	1
4	133	132	1
5	132	131	1
6	131	130	1
.	.	.	.
.	.	.	.
.	.	.	.
136	2	1	1

Table 1: the average reduction rate of the number of segments for consecutive iterations of the merge algorithms and the number of possible solutions at each stage. The process is repeated and yields a hierarchy of partition, with a stopping point being two segments.