AUTOMATIC MODEL ACQUISITION BY LEARNING *

Monika Sester
Institute of Photogrammetry
Stuttgart University
Keplerstraße 11, D-7000 Stuttgart 1
Commission III

Abstract

Recognition presumes having a model of what to recognize. This especially holds true for the recognition of objects in digital images. Such a model is usually formulated explicitly by humans. With the help of techniques from Machine Learning however, it is possible to automatically construct models from given examples.

The paper reviews several learning techniques and focuses on the automatic model construction with formal grammars. Both the requirements and the potential of such techniques are demonstrated with an application in the domain of landuse classification. A model for an agricultural parcel structure is used as one component in a system to recover land use maps from remotely sensed data.

Keywords: Artificial Intelligence, Machine Learning, Image Interpretation

1 INTRODUCTION

In order to derive landuse information from airborne images usually multispectral classification is used. But there is more than radiometric data in the images: both texture and geometry contain information about different landuse types. Each of these information sources needs a model to extract the relevant knowledge. This paper concentrates on the analysis of geometric models.

As shown by Janssen et al. [1991] using maps for classification greatly improves the classification result. When however no maps are available, a model for the parcel aggregation has to be provided. Such a model has to be a very general description, since it is impossible to represent any kind of possible parcel aggregation. Thus there is a request for a so-called generic model, where not only the object parameters but also the structure is free to a certain degree. An object parameter in the case of the parcel aggregation structure is e.g. the size of an individual parcel; the structure is reflecting the relations among the object parts (the number of neighbors of a parcel). „Polygon“ is a generic description for a parcel, in contrast to a list of n coordinates of the n points of the polygon.

Normally the models are formulated explicitly by humans. This is adequate as long as the objects are clearly definable and have distinct features. Often a collection of prototype objects is available, but still it is not clear in advance, which are the relevant parts of the object and which are its features and relations. In the terminology of knowledge representation the examples denote the *extensional* description of the objects. The task of Machine Learning techniques is to make this implicit knowledge explicit, thus end up in an *intensional* description.

In the paper the special problem of the parcel aggregation structure is analyzed and a strategy to extract a parcel model from examples is presented. The prerequisites for the automatic model acquisition are briefly sketched: given the examples, the internal structure of the data has to be extracted. The structure is revealed in a clustering process by grouping objects which are similar in some sense. The resulting graph is represented with the help of formal grammars, where each node is coded by a grammar rule. In order to give respect to the possible variety of the structure and also to noise effects, each node is now considered as the outcome of a random experiment. In a subsequent statistical analysis the node parameters are estimated, and so the statistic inherent in this structure is revealed.

The theoretical background thus lies on Machine Learning techniques, Knowledge Representation and Spatial Processes. After an overview of the project to which this contribution belongs, a short review of Machine Learning techniques is given, with special focus on model acquisition for Computer Vision purposes. The subsequent section is concerned with statistics. Finally an example for the model extraction from examples is given and the feasibility of this model is demonstrated.

2 KNOWLEDGE BASED LANDUSE CLASSIFICATION

Knowledge based image interpretation is performed by using any kind of information source available. A program to extract landuse information from aerial images can base on radiometric information in a traditional multispectral classification, but also on information about the geometry of objects. Thus not only mere grayvalues, but furthermore structural information about the objects is made use of. In order to integrate different sources of knowledge, the Minimum Description Length-principle (MDL) can be applied. MDL primarily allows to treat structural and numerical pieces of information within one compound process. In [Pan and Förstner 1992] a strategy for this task is presented.

Applying MDL presumes knowledge about the probability of the influencing factors. Knowing the whole functional chain

*This research is supported by the Deutsche Forschungsgemeinschaft in project SFB 228 „High Precision Navigation“

from grayvalues to geometry and structure, the probability for a hypothesized interpretation can be stated and evaluated:

$$\begin{aligned} P &= P(D,I,G,S) \\ &= P(D|I,G,S)\cdot P(I|G,S)\cdot P(G|S)\cdot P(S) \end{aligned} \tag{1}$$

or in terms of selfinformation or description length ($L = -\ln P$):

$$\begin{aligned} L &= L(D,I,G,S) \\ &= L(D|I,G,S) + L(I|G,S) + L(G|S) + L(S) \end{aligned} \tag{2}$$

In the formulas,

- S denotes the structural model description, the ideal geometry
- G describes the deviation of the real geometry from the ideal one
- I denotes radiometry and texture of the segmented image
- D corresponds to the original image data, describing signal, noise and outliers

Interpretations yielding the highest probability, or the shortest description length resp., are considered as the best ones. Evaluating an interpretation in terms of MDL presumes the functional dependencies of all the contributing knowledge sources to be modelled.

Pan and Förstner [1992] give a sketch of the use of the MDL principle for the interpretation of different landuse classes in airborne images: following an information preserving smoothing, in a second step segmentation techniques lead to edge and region information. A subsequent grouping process which is governed by the hypothesized model, namely polygonal areas, leads to a segmentation containing only polygonal boundaries. Still this representation is not complete and will contain ambiguous or false information. Basing on the model of the aggregation structure (S), hypotheses may be formulated. The best interpretation, the one yielding the shortest description, is found in a search process.

This paper is concerned with the last term of the formulas ((1) or (2)), namely the structural aspect. A functional and probabilistic description of the model structure is derived.

3 MACHINE LEARNING TECHNIQUES

Machine Learning is a branch of Artificial Intelligence which is of increasing interest in the AI community. Especially in the domain of knowledge acquisition for expert- or information systems there is a great demand for such methods. According to Simon [1984], learning denotes changes in a system to do the same task with the same input data more efficiently and effectively the next time. Michalski [1984] simply defines learning as a transformation of the representation. The new representation has to be „better" in some sense. In order to perform such a learning task, the notion of „better" has to be specified. The new representation ...

- ... mostly is not generated for its own sake, but is the basis for subsequent processes. In order to control and verify the new representation, it is given in a language, that is understandable by humans.
- ... supports and eases the handling of subsequent processes, like object classification, recognition, or location.
- ... is more compact than the old one: the task of „learning from examples" starts from a collection of examples and ends with a general description of these examples. The examples need more storage than the general description. Learning therefore supports data reduction.
- ... is explicit in contrast to the old one: knowledge acquisition often has to deal with 'diffuse' expert knowledge. Learning can structure this knowledge.
- ... is more general than the old one.
- ... can reveal new facts about the data.

Learning comprises three major considerations:

- the representation of the given data (input) and the desired data (output).
- a strategy to accomplish the transformation of the data from given extensional into the intensional representation.
- in order to evaluate the quality of the new representation and to distinguish different possible hypotheses an evaluation measure has to be given. This measure forms the basis to decide when the given aim of the learning has been reached or when to generate new hypotheses.

Following an historical sketch of Machine Learning research, the main techniques are shortly presented, along with exemplary programs from the domain of structural learning for Computer Vision purposes.

In the beginning of Machine Learning research, there was the wish for a general purpose learning system which starts without any initial structure or task dependent knowledge. Many of the early Neural Network approaches date to that phase. The limitations of these approaches, however and the idea of modelling human behaviour led to the development of programs which base on symbolic descriptions of the data. Representation schemes like logic, graphs, grammars were used to describe both features of the objects and relations. Since the mid 70 *ies* it is agreed upon, that learning does not start from scratch, but has to be incorporated in a broad knowledge framework. Thus task specific programs were developed, where the amount of background knowledge is manageable. This development reflects the change in data representation from numerical to structural.

Besides discerning Machine Learning techniques according to the knowledge representation into numerical and structural issues, a second distinction can be made considering whether the learning process is supervised or unsupervised. In supervised approaches, the training instances are presented along with a classification („learning from examples"), whereas unsupervised techniques automatically find a classification based on clustering or grouping processes („clustering"). Principally, clustering methods can only classify patterns, but do not give an explicit description of them. The result of a classification is just a distribution of the examples to different object classes, but not a description of the features of the classes. A subsequent characterization step (e.g. with learning-from-examples techniques) has to give a class description. Clustering techniques range from optimization methods with a given number of classes, over hierarchical issues resulting in binary classification trees called dendrograms, to clumping techniques which give clusterings in which object classes may overlap.

All learning systems try to generate possible solutions by grouping object features (possibly using additional background knowledge). This generation of the groupings in general is a search process which requests for suitable heuristics.

3.1 Numerical learning

This approach is chosen, when the numeric data is given in terms of feature- or attribute-vectors. These features span an n-dimensional space. The task of a classification process is to divide this feature space into several regions, the pattern classes [Niemann 1981]. Learning is reduced to an estimation of the unknown parameters, which link the observable features. For this task, a broad range of techniques is available.

In unsupervised numerical approaches (numerical taxonomy), grouping of the objects is performed on the basis of similarity. This similarity measure is the value of a numeric function applied to two objects (e.g. Euclidian distance). Thus objects which are most similar are grouped together, objects which are least similar are distinguished and form different object classes. Similarity measures can be both context-free (depending only on the attribute values) or context-sensitive, where the similarity is depending also on the value range of the attributes.

Neural Networks also deal with numerical data. Basically there are input and output patterns which are linked through one ore more layers by sets of weights. The learning task is to adjust these weights [Rumelhard and McClelland 1986].

3.2 Structural approaches

Conceptual clustering is a extension of numerical taxonomy to symbolic data. Conceptual clustering techniques typically do not only consider the objects' features and the context, but furthermore possible concepts or restrictions on the objects involved - so-called background knowledge - which may be used to describe the objects. Thus the similarity of two objects strongly depends on the quality of the concepts that describe the objects and their relations and possible limitations.

A classical concept to acquire models from examples was presented by Winston [1975]. His ARCH-program first derives a structural representation of the example objects in terms of a semantic network. In the following step the class descriptions of the different examples are constructed: the first positive example is hypothesized as the model, while the following positive and negative examples serve to correct or restrict the current model. In contrast to a complete object description, a model is generated, which is sufficient to distinguish an object from other objects in the knowledge base. Winston relies on noise free data; the resulting model is also represented in a semantic network.

A successor of Winstons program was developed by Connell and Brady [1985]. Their system is capable to handle real world data, namely real images. Their model generation strategy states, that if two objects are the same, then the differences between them should be irrelevant and can be deleted. The program expects only positive examples for an object class. Although handling with noise, it cannot treat outliers: these form a new object class.

Wong and You [1985] present a program starting with examples in an attributed graph. In an estimation process, the relevant attributes and the relevant relations, along with the corresponding probabilities are gained. This structure, which represents the model, is called a so-called random graph.

Segen [1988] developed a program to learn descriptions for 2D-objects from examples. In contrast to most other learning techniques, his program is not based on a fixed set of attributes and relations, but is able to generate its own descriptors. The program starts with the objects (given as object contours) and calculates points with curvature maxima, which are denoted descriptors of the first level. Then it generates new descriptors by successively grouping the features of the previous level. The resulting hierarchical graph is analyzed statistically. The system can treat noisy data. The drawback is that the resulting representation bases on the descriptors chosen, which are not interpretable by humans.

Also Stier [1991] starts from the idea, that object representations relying on a predefined set of descriptors can only be as expressive as the descriptors are. Therefore he argues for a system that is capable of deriving its own appropriate descriptors. He presents a learning technique which starts from elementary, general knowledge about objects, represented in logical assertions. In the evaluation phase, a general object is hypothesized and the rules from the knowledge base are applied to it. The examples serve to verify (accepted or rejected) these hypothesized new rules. The evaluation is performed in an exhaustive search manner, where the matching criterion is the exact fit with the examples. He demonstrates his strategy on simple examples. Starting with elementary knowledge about polygons (parallelity, straightness, ...) he learns the concepts of special types of polygons (triangles, squares, rectangles).

All structural learning techniques presented have in common, that they rely on a hierarchical representation of the objects. In a first step, the structural description of the examples is derived. Then this description is generalized to a model. For both steps various strategies may be used. These however depend strongly on the data and the task, and no general technique is available up to now.

4 DATA REPRESENTATION WITH FORMAL GRAMMARS

Thinking of patterns in terms of sentences makes it possible to apply techniques from formal language theory to pattern recognition [Fu 1982]. With the help of Formal Grammars the knowledge about the structure of observations is represented in symbolic form. A grammar consists of the tuple

$$G = (S, V_N, V_T, P)$$

where S denotes a Startsymbol, V_T and V_N the Terminal and Nonterminal vocabulary (symbols), and P a set of rules (production rules) describing, which symbols may be replaced by other symbols. Given a startsymbol, any structure can be derived within the domain of the grammar, simply by replacing the nonterminal symbols with the help of the production rules. This procedure stops, i.e. a structure is generated, when only terminal symbols (which cannot be replaced) occur in the sentence.

In general, formal grammars allow both for generation of new structures and for deciding, if an unknown structure is explainable with the given grammar [Cohen and Feigenbaum 1982].

Formal grammars rely on noise and error-free data. However, when real physical processes are involved, the grammar has to cope with non perfect data. To this end the concept of formal languages is extended to a stochastic grammar, where each production rule is assigned a probability of occurrence.

With an attributed grammar, functional dependencies of the nonterminal symbols of the rules can be coded in a compact way.

5 SPATIAL PROCESSES

Spatial data processing deals with the analysis of spatially distributed patterns. The task is to find regularities among the data or to make assumptions on the underlying mechanism that generated the pattern [Ripley 1981].

A frequently applied model are stochastic processes, especially Poisson processes. In the scope of this paper, so-called Renewal Processes are of importance. Primal assumption of Renewal Processes is that a random experiment is repeated with the same assumptions and probabilities as the first experiment. Thus with the repetition the process really starts from anew. The model of a Renewal Process is usually applied in the analysis of defects of machine parts. Such parts may break down now and then. A break-down at one time instance does not affect the next defect, a feature denoted as the „lack of memory"-property. The probability of a defect itself is distributed with certain parameters (usually Poisson: λ). The probability of an event at time instance j is given by the following formula:

$$P(X=j) \;=\; \frac{\lambda^j}{j!}\exp^{-\lambda} \; ; \; E(X)=\lambda \qquad (3)$$

Instead of discrete time instances, also discrete spatial parameters can be modelled with this process.

6 AUTOMATIC ACQUISITION OF PARCEL STRUCTURE

The visible regularities in the agricultural parcel structure are due to fact that the subdivision was ruled by certain criteria: the parcels are of a reasonable size and of a simple form (e.g. rectangles with one „long" side), in order to be manageable with machines. Possible other aspects like history, sociology or aesthetics will not be considered in this context, as only observable features are taken into account.

In spite of the underlying planning, the aggregation structure is not unique, a model is not easy to determine. At a first glimpse, a simple model could be a collection of parcels, each of which is represented by a polygonal boundary. This representation however contains no information about the relations of the parcels. Without specifying relations, a parcel structure would just be a random parcel puzzle, neglecting the neighborhood relations which are quite obvious: most parcels are connected to at least one other parcel of similar form and size, sharing one common border completely. Thus a more elaborate model is required, taking the structure of the object into account.
In maps or images there is a lot of exemplary data available, thus the idea is to use structural and statistical learning techniques to automatically derive a parcel model from examples.

The task is to turn specific knowledge (examples) into general one (model). In the spirit of Winston's approach, first a structural description of the examples is generated, then this structure is generalized to a model. The resulting model description should intuitively fit the description humans have. This can be verified by using the model to generate new objects, i.e. simulate the generation process and compare the outcome with real data. In the following, the automatic acquisition of a generic parcel model is sketched.

The program starts with the observable information: an example of a parcel aggregation represented as line segments (see Figure 1).

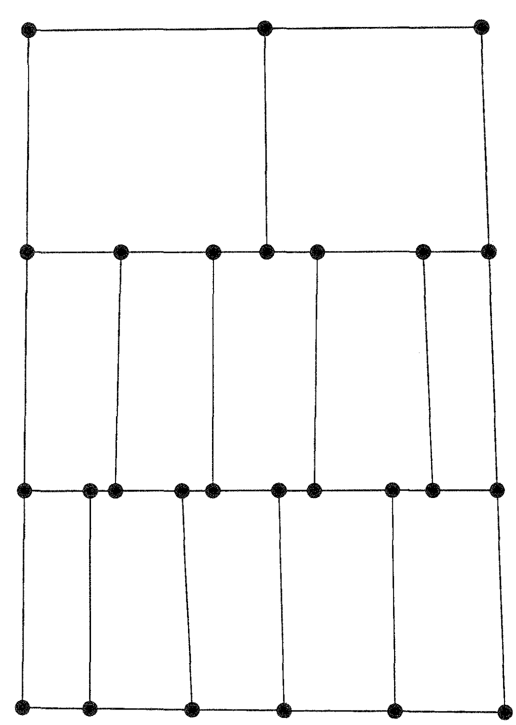


Figure 1: Input data: digitized line segments

The structure, namely the individual parcels and the relations between them, has to be extracted with the help of a learning strategy. A clustering process is applied, which is grouping parcels which are similar in the sense of neighborhood. A iterative grouping leads to a graph, where the leaves represent the individual parcels, while the nodes stand for groups of neighbored parcels. The top node finally is the

„father“-area, all the others were derived from. This graph - which in this case is reduced to a tree - mirrors the generation process. The nodes in the tree form the structure elements of the model. Each node has certain attributes (like form and size) and certain relations to other nodes (the fields it is divided into). Thus the node information states, how a parcel with certain attributes is divided into smaller ones.

The nodes are considered as an outcome of a random experiment. An estimation procedure is applied in order to gain the structuring parameters (i.e. the parameters of the parcelling) and their probabilities. In that way the variability of the structure is evaluated with the help of a statistical analysis. Since the parcelling is relating spatial entities, the evaluation makes use of statistical spatial processes: the subdivision of a bigger parcel into smaller ones is modelled with a Renewal Process, i.e. the partitions are distributed with a common parameter of the Poisson distribution λ and the individual cuts are independent of each other. This modelling is motivated by the fact, that the size of an individual parcel is determined by factors that cannot be estimated from by visual knowledge alone, but mainly depends on legal aspects, namely the claims of its owner. In that way the sizes of the parcels can be considered independent.

Thus the generation tree reveals that parcelling is a recursive process: new parcels originate by dividing a big parcel into smaller ones. This recursive structure favours the representation scheme of formal grammars, where the model information is coded with an attributed stochastic grammar.

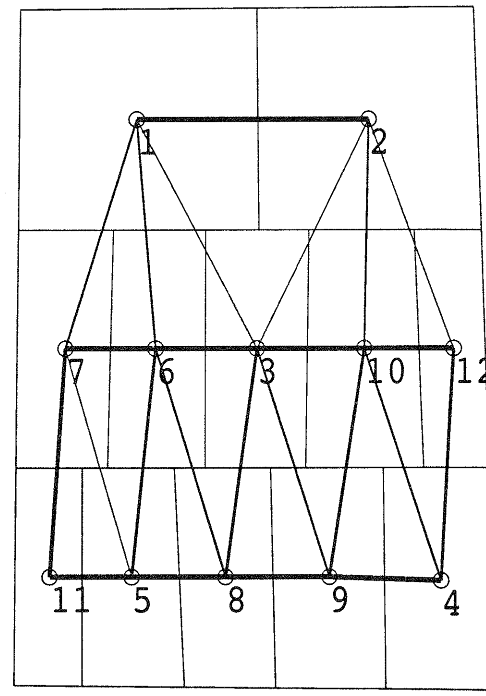


Figure 2: Region-Adjacency Graph: thick lines denote close relation

6.1 Structural analysis

The clustering process is a mixture between structural and parametric approaches. First the structure of the data is gained, then numerical values expressing the relations between the object parts are calculated in order to guide the clustering process.

Starting with line segments, in a first step the individual parcels are extracted by looking for a trace of segments forming closed contours. To this end the list of lines is cycled twice: the points are traversed clockwise for the outer-boundary of a region and counter-clockwise for an inner boundary.

In order to cluster the parcels a measure of similarity or connectedness has to be given. In this case the measure is defined by the adjacency of the parcels. Only parcels which are neighbored can be grouped together. Furthermore, the more complete the common border between two parcels is, the more „similar“ the parcels are, the closer they are together. In a region-adjacency graph the neighbored parcels (see Figure 2) are shown, where the degree of similarity is visualized by the thickness of the connecting lines.

This graph structure is subjected to a clustering procedure where similar regions (in sense just defined) are merged. A simple hierarchical clustering method is successively grouping parcels which share one common border. In that way a dendrogram is produced with the single parcels as leaves and the „father“-parcel at the top. Nodes in between form subparcels, which are further divided (see Figure 3). This tree reflects the generation process of the individual parcels: starting from a big area and dividing it successively.

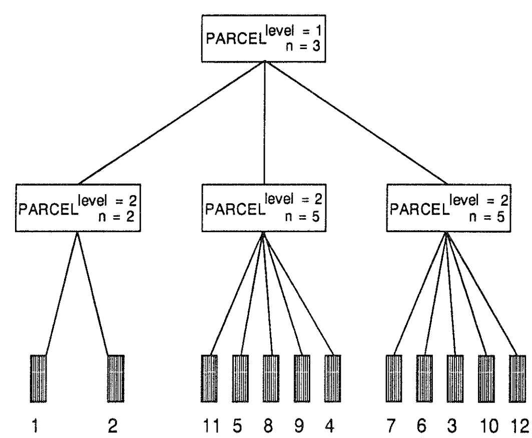


Figure 3: Dendrogram of grouped parcels

The information contained in these nodes (the nodes' attributes) are the size and the form of the parcel, the direction of the partition, the number of its successors and the number of previous partitions. This information is analyzed statistically in the next step.

6.2 Statistical analysis

Up to now the analysis produced a production rule for each individual parcel. The desired information is a set of stochastic production rules which state, how individual parcels are

divided. This division is dependent on the parcels size, form, and the number of partitions this parcel was subjected to already. With a statistical analysis an estimation of these dependencies and parameters, accompanied with an estimation of the probability of each rule is gained.
Global parameters like maximal and minimal parcel size, maximal and minimal ratio of shorter side of a rectangle to larger side, and maximal hierarchy level can be estimated in a first step. These parameters determine when a subdivision stops.

The dependency of the partition on the parcel attributes is modelled in terms of a Renewal Process.

In this paper, only the outline of the estimation process is given. In order to reliably estimate the model parameters, a big sample of exemplary data has to be evaluated. Furthermore, the functional dependencies of the different parameters have to be examined carefully. For simplicity in the following example dependencies on the size and form of the parcel and on the direction of the partition are neglected, and the following simplification is made:

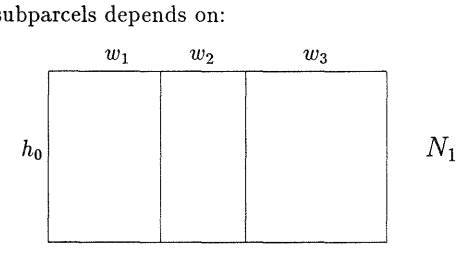
- Parcels are divided along their longer side.
- The partition of a parcel only depends on the hierarchy level, namely the number of previous partitions. This is a fact that is easily verified with visual inspection of the data: the first cuts of a parcel try to generate a few big parcels, while the following split these parcels into the final individual fields, by higher number of cuts. The parameter λ in equation 3 corresponds to the number of successors n. Different parameters λ are estimated depending on the hierarchy level.

The result of the estimation are both the functional dependencies and the corresponding probabilities. These values for the relation between hierarchy level *level* and number of cuts n for the above example (see Figure 3) are given in the following table:

level	1	2	
n	3	5	2
$P(n\|level)$	1.0	0.67	0.33

These values are gained by simply counting the possible numbers of successors of a level. On level 2 e.g. there are 3 parcel to divide. Two of them are cut into 5, one is cut into 2 parcels. Thus result the probabilities of 2/3 and 1/3 resp..

In order to rate an existing aggregation structure, the probabilities of the individual steps of its generation have to be evaluated in common. To this end the whole chain of dependencies has to be formulated. The probability of partition $P(part)$ of a parcel N_1 with parcel sides w_0 and h_0 into n subparcels depends on:



- $P(N_1)$: the probability of „father"-parcel N_1
- $P(n|level)$: the probability of n cuts given hierarchy level *level*
- the probability of dividing parcel N_1 into n cuts with widths w_i

$$P(w_1, w_2, \cdots, w_n|N_1, n) = P(w_1|N_1, n) \cdot \\ \cdot P(w_2|w_1, N_1, n) \cdots \\ \cdot P(w_n|w_1, w_2, \cdots, w_{n-1}, N_1, n)$$

  Due to the „lack of memory"-property of the Renewal Process, the probabilities of the individual cuts w_i can be considered independent from each other, i.e. depending only on the „father"-parcel and the number of previous partitions n, resulting in

$$P(w_1, w_2, \cdots, w_n|N_1, n) = P(w_1|N_1, n) \cdot \\ \cdot P(w_2|N_1, n) \cdots P(w_n|N_1, n)$$

- the probability of the width w_i of a parcel given the number of partitions n is computed with equation 3:

$$P(w_i|N_1, n) = P(X = \frac{w_0}{w_i}|\lambda = n)$$

The probability of partition $P(part)$ is then:

$$P(part) = P(N_1) \cdot P(n|level) \cdot \\ \cdot P(w_1|N_1, n) \cdot P(w_2|N_1, n) \cdots P(w_n|N_1, n)$$

All the other dependencies have to be considered correspondingly.

6.3 Representation in attributed stochastic grammar

After estimating the model parameters, the stochastic grammar can be set up.

- Vocabulary:

$$\begin{aligned} V_N &= PARCEL_A \\ V_T &= parcel_A, relation \\ relation &= |, - \end{aligned}$$

- Startsymbol $S = PARCEL_A$
- Production rules P:

$$\begin{aligned} PARCEL_A &\overset{P(part)}{\longrightarrow} PARCEL_A relation PARCEL_A \\ PARCEL_A &\overset{P(stop)}{\longrightarrow} parcel_A \end{aligned}$$

Nonterminal symbols (denoted in uppercase letters) stand for intermediate parcels; terminal symbols (lowercase letters) give the final parcels or the spatial relations between the parcels resp. (| = left-right; $-$ = top-bottom connection). Each parcel has corresponding attributes A.
This grammar represents a compact model description. It can be used in an algorithmic way to produce new objects, i.e. new parcel aggregations. Given a startparcel, new parcel are generated by applying the rules of the grammar. An example is given in the next subsection. On the other hand, the grammar can be used to object recognition: if an unknown object can be explained by the grammar, then it is recognized.

6.4 Parcel generation with model

In order to evaluate the model visually, new parcels are generated with the help of the model. Given a bigger parcel, it is divided into smaller ones with rules of the grammar. Starting at hierarchy level 1, the number of subparcels is taken from the functional relation table: $n = n(level = 1) = 3$; the associated probability is $P(3|1) = 1.0$. Thus the parcel is cut into three subparcels. Formally this results in a structure:

$$PARCEL \overset{P(part)}{\longrightarrow} PARCEL - PARCEL - PARCEL$$

The production rules are now applied to the subparcels correspondingly. Being in level 2 now, the number of successors is either 5 (with probability 0.67) or 2 (with lower probability 0.33). In the example, the color indicates the probability of the partition: bright parcels have high, dark ones low probability. The first partition (Figure 4) - which shows clear similarities to the example (see Figure 1), is more probable than the second one (Figure 5).

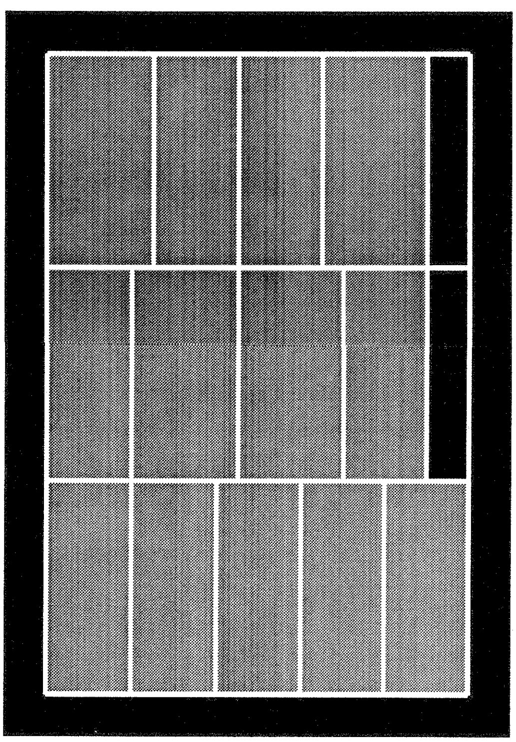


Figure 4: Aggregation structure generated by model - high probability

7 FINAL REMARKS

A prerequisite for any interpretation of images is to have adequate model descriptions for the expected objects in the images. As these are normally quite complex, there is a need for automatic extraction or generation of such models. In general, although examples for an object can be identified and enumerated, they can not be described in a compact form. Learning techniques are a means to solve this problem: they allow for making a interior structure (hidden in the examples) explicit. Which technique to chose, however, is dependent on the type of object.

The approach to automatic model acquisition from examples given in the paper basically is a general one, since it generates a structural representation of the data. Such descriptions, specifying an object in terms of parts and relationships are useful for high level image interpretation tasks, dealing with complex real world data.

The result of the modelling is not only a description of the object, but the model also reflects the statistic in it. This opens the way to a compound analysis of data of different knowledge sources basing on the MDL principle.

The paper showed that it is possible to extract a representation for parcel structures from examples. The new representation gained from the original line representation is more compact, of generic type and thus more suitable for subsequent object location or recognition tasks.

LITERATURE

Cohen, P.R. and E.A. Feigenbaum [1982]: The Handbook of Artificial Intelligence, Vol. 3. Kaufmann Inc., Los Altos, California, 1982.

Connell, J.H. and M. Brady [1985]: Learning Shape Descriptions. *Proceedings of the IJCAI*, pp. 922-925, 1985.

Fu, K.-S. [1982]: Syntactic Pattern Recognition and Applications. Prentice Hall, New Jersey, 1982.

Janssen L.L.F and H. Middelkoop [1991]: Knowledge Based Image Classification; the use of knowledge about crop rotations in a Thematic Mapper Image Classification Procedure. Report 40, The Winand Staring Centre, Wageningen, The Netherlands, 1991.

Michalski, R.S. [1984]: A Theory and Methodology of Inductive Learning. In: *Machine Learning*, Michalski, R.S. et al. (ed.), Springer-Verlag, Berlin, 1984.

Niemann, H. [1981]: Pattern Analysis. Springer Verlag, Berlin, 1981.

Pan, H.-P. and W. Förstner [1992]: Stochastic Polygon Map Grammars: A Generic Model for Understanding Landuse Maps and Images in Remote Sensing. In: *Robust Computer Vision*, Wichmann, 1992.

Ripley, B.D. [1981]: Spatial Statistics. Wiley, New York, 1981.

Rumelhard, D.E. and J.L. McClelland [1986]: Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, Cambridge, 1986.

Segen, J. [1988]: Learning Graph Models of Shape. *Proceedings 5th International Conference on Machine Learning*, pp. 29-35, 1988.

Simon, H.A. [1984]: Why should Machines learn ? In: *Machine Learning*, Michalski, R.S. et al. (ed.), Springer-Verlag, Berlin, 1984.

Stier, S. [1991]: Bildverstehen als Lernprozeß aus elementarem Wissen, PhD thesis, Universität Hamburg, 1991.

Winston, P.H. [1975]: Learning Structural Descriptions from Examples. In: *The Psychology of Computer Vision*, P.H. Winston (ed.), McGraw-Hill, New York, 1975, pp. 157-209.

Wong, A.K.C. and M. You [1985]: Entropy and Distance of Random Graphs with Application to Structural Pattern Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 5, pp. 599-609, 1985.

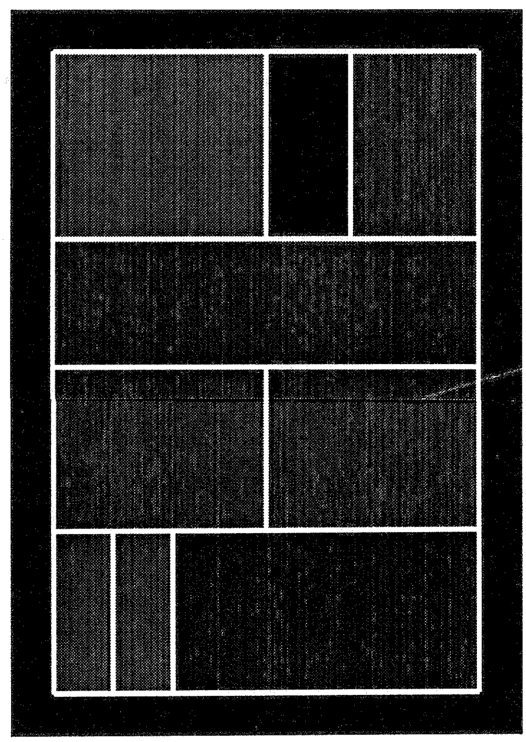


Figure 5: Aggregation structure generated by model - low probability