Photogrammetric Record, 16(96): 963–985 (October 2000)

IMAGE ANALYSIS FOR GIS DATA ACQUISITION

By C. HEIPKE, K. PAKZAD and B.-M. STRAUB
University of Hanover, Germany

(Paper read at a Technical Meeting of the Photogrammetric Society on 14th March, 2000)

Abstract

The automatic interpretation of aerial and satellite imagery by means of image analysis is currently one of the major research tasks in photogrammetry and related disciplines. The primary goal is the automatic extraction of visible, spatial topographic objects from imagery. The extracted objects represent one possible form of input for creating geographic databases. In this paper, different aspects of image analysis are discussed and a framework is provided for scene interpretation, which is based on the integration of image analysis and a GIS data model. Two examples concerned with the combined extraction of roads and trees, and with the multitemporal interpretation and monitoring of moorland, are given to illustrate the research.

KEY WORDS: automation, data capture, GIS, image analysis, scene modelling

INTRODUCTION

GEOGRAPHICAL INFORMATION SYSTEMS (GISs) have received major attention over recent years. A particularly important challenge is the task of populating the GIS databases with geo-objects. In order to be useful in various applications, the geo-objects, also referred to as GIS data, need to be of high quality in terms of currency, completeness, homogeneity, geometric accuracy and semantic correctness. Photogrammetry and remote sensing have proved their ability to meet these requirements (Englisch and Heipke, 1998) and therefore provide the primary technology for GIS data acquisition and update.

The increasing coherence between the acquisition and the further use of the data has had consequences for the relationship of photogrammetry and remote sensing on the one hand, and GIS on the other. In the past, these were distinct disciplines, being mainly connected through data transfer from imagery to the GIS data base. Today, besides a bidirectional link needed to use existing GIS data as prior information for updating, a trend for complete integration can be observed. In this sense, photogrammetry and remote sensing can be described as a three dimensional data acquisition module of GIS, using multisensor, multispectral, and multitemporal images, including data from laser scanners and interferometric synthetic aperture

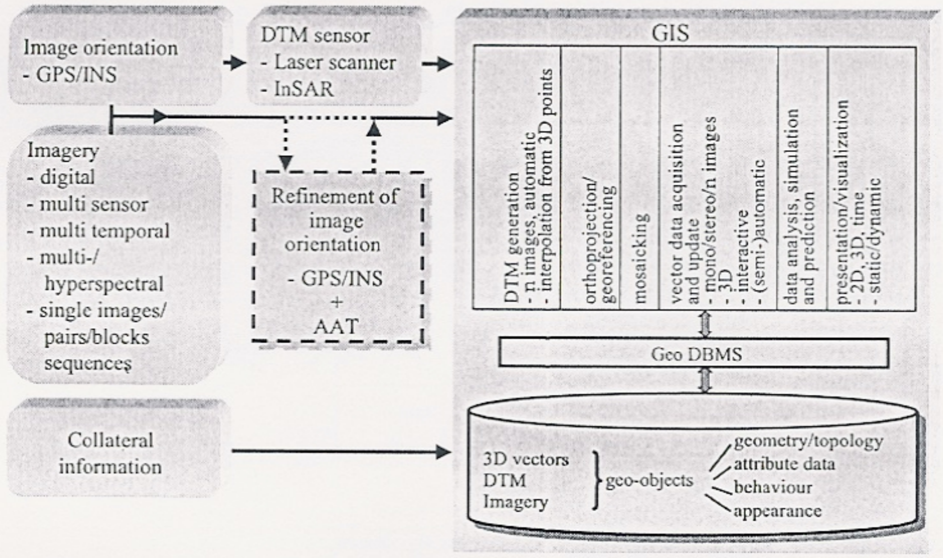

FIG. 1. Conceptual integration of photogrammetry, remote sensing and GIS.

radar (InSAR) as primary data sources. For orientation purposes, the corresponding sensor system is equipped with a GPS (global positioning system) receiver, an IMU (inertial measurement unit) and software for AAT (automatic aerial triangulation); collateral information, such as the co-ordinates of ground control points (GCPs), is also needed. In essence, the results of the sensor system are geo-coded images, available immediately after data acquisition. All tasks connected with further data processing can then be considered as GIS modules working on a common database containing the geo-objects (Fig. 1). The geo-objects are described in a feature class catalogue or GIS data model in terms of geometry, topology, attributes and radiometry (necessary for visualization and automatic object extraction from the images). GIS modules for data processing comprise the generation of digital terrain models (DTMs), ortho-images and ortho-image maps, the acquisition of vector data, and also the analysis and visualization of the data. First steps towards realizing such a GIS have been undertaken by the major competitors in the field, such as ESRI in co-operation with ERDAS, LH Systems in co-operation with Laser Scan, and Z/I Imaging.

It is well known that the GIS data constitute the most valuable part of any GIS, partly because of the high cost involved in data acquisition and update. Therefore, major research efforts have been concentrated on partly, at least, automating the data processing steps that have been mentioned. In terms of progress to date, automatic image orientation and DTM generation are operational to varying degrees, but also face increasing competition from techniques such as GPS/IMU, laser scanning and InSAR where the required information is measured directly rather than being determined indirectly from images. Orthoprojection and mosaicking have already made major inroads into practice and must be considered as the industry standard. The automatic extraction of vector data from images (also called image analysis), however, is still mainly a topic for research; only multispectral classification is being used in practice. It is now appreciated that image analysis is a highly challenging field and its complexity is often underestimated because humans are so much at ease with this task.

Image analysis forms the main focus of the research described. One of the main research tasks in photogrammetry and computer vision is image analysis for topographic mapping; Ebner *et al.* (1999) and Förstner *et al.* (1999) give an overview of its current status. This paper commences by stating some basic considerations relevant to image analysis, followed by discussion of methods for reducing the complexity of image analysis by using multiple image scales and related aggregation levels of a GIS data model. Two examples of the authors' work are then given. The first describes the combined extraction of roads and trees from colour infra-red (CIR) imagery, demonstrating the advantages of modelling more than one object class in isolation. The second example is concerned with the interpretation and monitoring of moorland from black and white and from CIR images. The emphasis of this example is on multitemporal image analysis. Finally, some remarks are made concerning future research.

SOME BACKGROUND IN IMAGE ANALYSIS

Models, Formal Knowledge Representation and Processing Strategy

About two decades ago, Rosenfeld (1982) defined image analysis as "the automatic derivation of an explicit meaningful description of physical objects in the real world from images". It is clear that in order to be able to recognize an object in an image, the object itself first has to be described in the form of some kind of model. This model needs to capture the particular characteristics and the invariant aspects of the object in order to be distinguishable from other objects. Besides a geometric description containing information about size, shape and so on for the object, radiometric information (in other words the appearance of the object in images in the form of brightness values, texture and spectral characteristics) must be included in the object model.

In sympathy with developments in knowledge representation in artificial intelligence, most image analysis approaches today are based on a clear separation between an explicit description of the object model and the actual process of object extraction. The object model is stored in a so-called knowledge base. As well as the object model, a model for connecting object and image space is introduced. In the case of aerial imagery, the obvious choice is a perspective projection. Approaches designed along these lines are often referred to as "model-based" or "knowledge-based" image analysis in order to emphasize the role and the explicit representation of the object model. Nevertheless, these terms are a little misleading, because any method trying to automatically extract objects from images obviously relies on some sort of model and knowledge.

In order to use this knowledge within a computer system, it must be formally represented and then translated into some computer-compatible form. In topographic applications of image analysis, the formal knowledge representation is often based on semantic nets (Niemann *et al.*, 1990; Sagerer and Niemann, 1997). These nets are directed acyclic graphs. They consist of nodes and edges in between the nodes. The nodes of the semantic net model the objects of the scene and their representation in the image. Two classes of nodes are distinguished: the *concepts* are generic models of the object and the *instances* are realizations of their corresponding concepts in the scene observed. Thus, the knowledge base which is defined prior to image analysis is composed of concepts assembled in a concept net. During the interpretation, a symbolic scene description is generated consisting of instances which form the instance net. The relationships between the objects are described by edges or links of the semantic net. Normally, only a limited number of different relationships are used in semantic nets. The *is-a* relation describes the specialization of objects, introducing the property of inheritance. Along the *is-a* link, the description of the parent node is inherited to the more specialized node. This description can, however, be overwritten locally. The second important relationship is the *part-of* link, which offers the possibility of describing an object by means of its parts. Thus, the detection of an object can be simplified to the detection of its parts. Finally, the transformation of an abstract description into its more concrete representation in the data is modelled by the *concrete-of* relationship, abbreviated to *con-of*. This relationship allows for structuring the knowledge in different conceptual layers, for example a scene layer and an image layer.

In order to make use of the knowledge represented in the semantic net during image analysis, a processing strategy is needed. The role of this strategy is to define how and in which order the image interpretation proceeds. In model-based image analysis, it is usual to pursue a top–down approach. Starting at the top of the semantic net containing the complete scene of the landscape, a concept is instantiated if, and only if, all of its parts have already been instantiated. Thus, one proceeds downwards in the net, until the required evidence of some object part is found in the image to be analysed. The sequence in which the concepts are processed often depends on a set of rules, which essentially state that those parts which can be extracted easily, reliably, and fast are to be treated first in order to reduce the search space for the further analysis. As such, the processing strategy can be implemented in the form of a tree search such as the A* algorithm.

Most of the approaches for topographic applications existing today were designed for the extraction of one object class only, for example roads (Baumgartner *et al.*, 1997), buildings (Fischer *et al.*, 1998) or vegetation (Borgefors *et al.*, 1999), partly due to the high complexity of the real world and the corresponding images. These approaches have reached a comparatively high level of success in terms of research results. However, the separation of image analysis into different subtasks leads to restrictions in more complex imagery, in which different object classes appear close together and disturb each other. Therefore, in the approaches mentioned above, it must be ensured that the image to be analysed actually contains objects of the required class, and that they are well separated from other non-modelled objects. Another deficiency of existing methods is that they only take into account a single epoch in time. This limitation becomes particularly important with regard to objects changing over time, such as vegetation. In order to overcome these deficiencies, for scene models containing different objects and images taken at different epochs, their change over time must be considered.

Models in Image Analysis and GIS Data Models

The object model described is in many ways similar to the GIS data model mentioned above, which is not surprising because both models describe the same objects. However, the model for image analysis must also include the appearance of objects and disturbances such as occlusions and shadows. Therefore, it must inherently be three dimensional. It should also contain the relationships with surrounding objects, because the context given by these is often essential for the interpretation. For instance, an object which is difficult to identify but which is located on a road, is more likely to be a car than a hut. Thus, the context "road" is advantageous for the interpretation of cars. The GIS data model, on the other hand, often contains aspects which are not visible in an image, for example object names and legal boundaries. Obviously the more similar the two models are, the easier an exchange of the data becomes and also the easier an integration of image analysis and GIS becomes. Since

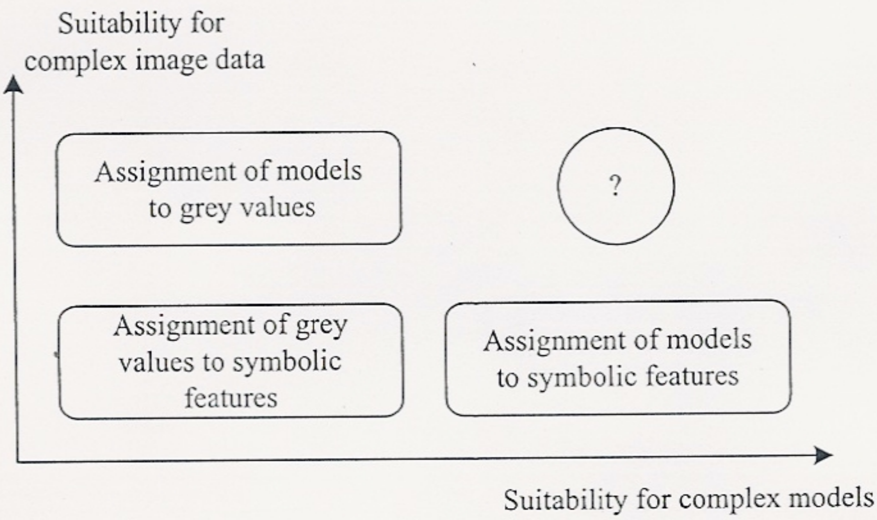

FIG. 2. Solution space of different methods for image analysis.

the GIS data model is usually dictated by the application, it is advantageous to design the object extraction model as closely related to the GIS data model as possible. In doing so, it is possible to make use of the hierarchical nature of most GIS data models. First, the landscape is subdivided into four superclasses: *settlement*, *water*, *forest* and *open landscape*. In many GIS data models, quite similar classes can be found as the highest level of abstraction in a hierarchically structured data model. These superclasses can than be used as global context knowledge about the observed subset of the real world. Global context knowledge is related to imagery at a coarser ground resolution, which leads to an abstraction of the image content (Lindeberg, 1994). The global context defines the frame for the extraction of individual objects and makes the automatic interpretation of higher resolution imagery more feasible (Rapp, 1995).

Reduction of Complexity: a Key for Successful Image Analysis

Following the proposition made by Suetens *et al.* (1992), and also used by Mayer (1998), the existing methods of image analysis can be mapped into a two dimensional solution space (Fig. 2). One axis describes the suitability for complex models and the other axis the suitability for complex image data. As can be seen, methods for processing either simple images or simple models are available. However, complex models in conjunction with complex image data cannot be handled appropriately with existing methods, as indicated by the "?" symbol in Fig. 2. Thus, a reduction in complexity is necessary. Since the model complexity is usually dictated by the application, only the image complexity can be manipulated. It can be reduced by means of scale-space transformation (Mayer, 1998) and by global context knowledge, that is by restricting the image area to be processed.

In order to solve a problem with complex models and complex images, it can be argued that the following strategy is feasible.

(i) Reduce the image complexity using a scale-space transformation and simultaneously reduce the model complexity for processing the coarse-resolution imagery. Multispectral classification is an example of a simple model.

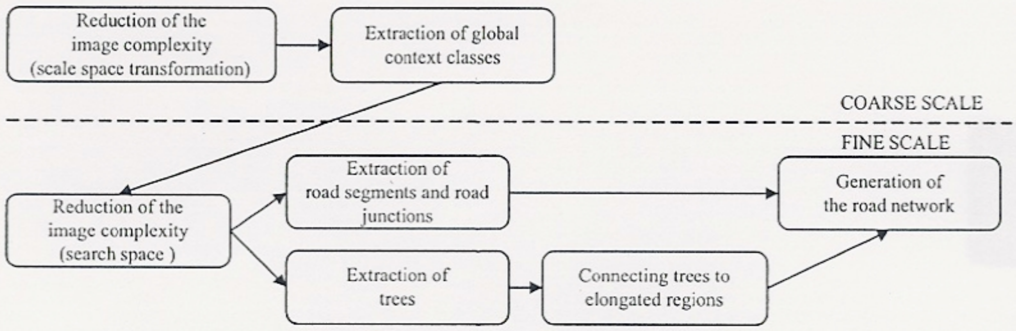

FIG. 3. Strategy for interpretation based on global and local context in different scales.

(ii) Reduce the image complexity by using global context knowledge in the form of the results of (i) for processing the higher-resolution images. As an alternative to employing the results of step (i), existing GIS data can be used as global context for complexity reduction.

An early example of implementation of this strategy is the multiscale road extraction approach adopted by Baumgartner *et al.* (1997). Road extraction is only performed in those global context areas where it is known to work reliably. Further analysis then depends on local context. The strategy proposed above is also employed in the applications described in the following part of this paper.

COMBINED EXTRACTION OF ROADS AND TREES USING GLOBAL AND LOCAL CONTEXT

The combined extraction of roads and trees from CIR imagery is taken as the first example. In order to remain independent of existing GIS data, a method is proposed which is based on imagery alone, using a multiscale approach consisting of an enhanced multispectral classification at a coarse level, followed by object extraction with a fine level of image resolution. The strategy presented by Baumgartner *et al.* (1997) is enhanced by means of explicit modelling of local context knowledge which is given by trees, and especially by rows of trees occluding a road. In order to be able to use this knowledge, not only roads but also trees and rows of trees are automatically extracted. The general procedure for this approach is depicted in Fig. 3 and is explained in more detail below.

Description of the Input Data

The images used in this study were acquired in summer 1997 using the digital photogrammetric assembly (DPA) sensor. They have a ground resolution of approximately 0·8 m. Reduced resolution images were computed in morphological scale space with 1·5 m and 12 m pixel sizes on the ground. The DPA camera operates in four spectral bands. Three bands (red, green, blue) are in the visible range of the spectrum and the fourth lies in the near infra-red.

The hierarchical structure of the data model of the German Authoritative Topographic-Cartographic Information System (ATKIS) was used as the basis for the data model. The landscape is subdivided into different object classes which are represented in different aggregation levels. Both the object classes and the aggregation levels are similar to those defined in ATKIS.

The size of the scene investigated is about 35 km^2. It is thus large enough to contain one or more villages, forest and agricultural land, as well as parts of the road network. The presence of a variety of topographic objects, which in Germany and Central Europe can often be found within a few square kilometres, is a prerequisite for the approach described.

Extraction of Global Context

As mentioned previously, global context is described by four superclasses: settlement, water, forest, and open landscape. This simple description of the landscape is in contrast to the relatively complex content of the images at the original resolution. In order to overcome this contradiction, a transformation in morphological scale space is applied to simplify the image content. The scale-space transformation can be regarded as an abstraction of the image content; for example, single trees in forest areas are aggregated to homogeneous regions (Heipke and Straub, 1999). The morphological scale space is related to the size of the objects and, therefore, it is relatively independent of sensor characteristics and the actual lighting conditions of the scene.

Global context knowledge was extracted from the image of 12 m resolution by an enhanced multispectral classification. The training areas for the multispectral classification are derived automatically. Based on the Normalized Difference Vegetation Index (NDVI) and the image texture (the local variance is used in this case), settlement and water areas were identified; both have low NDVI values, but water has a small variance value, whereas settlement has a high variance value. The remaining areas were subdivided into three classes by means of an ISODATA cluster analysis (Richards and Jia, 1999). The darkest region is assumed to contain forest and the two lighter regions are considered to be open landscape. Next, the largest regions in each of the five classes were selected and used as training areas for a multispectral classification. Subsequently, geometric constraints based on the object description in the GIS data model, namely size and compactness, were applied to refine the results. The examples of the object classes *settlement* and *forest*, originating from the multispectral classification, were filtered by means of these geometric constraints. All rejected areas, for example roads in open landscape which possess spectral characteristics similar to settlement but are not compact, and groups of trees and bushes which have the same spectral signature as forest but are too small, were labelled as open landscape.

Object Models and the Application of Local Context Knowledge

In this section, the models and strategies for the extraction of roads and trees are described, operating within the global context areas described above. The model for the extraction of trees is based on the object description in the GIS data model given. Road extraction was performed using an algorithm developed at the Technical University, Munich (Wiedemann, 1999). The reasons for the selection of these two object classes were: (1) in the context *open landscape* there is an obvious relationship between roads and trees, namely that "trees may occlude roads"; and (2) a further application of global context knowledge can be demonstrated, "the colour of trees is given by the colour of forest".

As mentioned, the model for trees is taken from the object description in the GIS data model. In the ATKIS feature class catalogue, the following constraints are found for the feature class 4201 *Row of Trees*. The data capture criterion for this object, according to ATKIS, is: "A row of trees has to be captured if it is longer than 200 m and lies near roads or is formative for the landscape". The maximum distance between two trees is not defined; the diameter of a single tree was used for this research. Single trees are not modelled in the ATKIS feature class catalogue, therefore a model of a generic individual tree has to be added. A simple model is

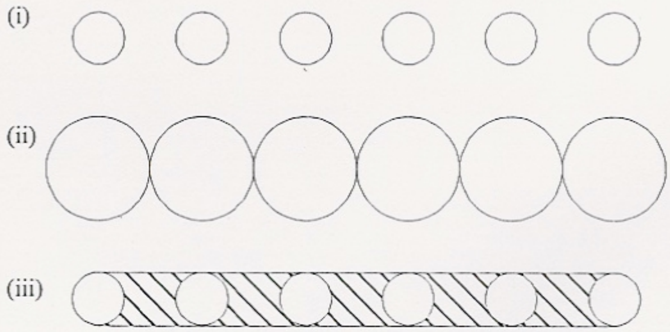

FIG. 4. Model employed to identify a row of trees.

sufficient for this task due to the restricted image resolution; more detailed tree models have been suggested in the literature, for example the detection of local brightness maxima (Pinz, 1989) or a rule based approach for the delineation of crowns (Gougeon, 1995). The authors used a round blob with a diameter of about 10 m and a characteristic colour to represent a single tree. The extraction of instances of the object *Row of Trees* in image space is executed as follows (Fig. 4). First, blobs with the specified size and the spectral appearance of forest are searched for inside the open landscape (Fig.4(i)). These regions are dilated in order to span the distance between single trees (Fig.4(ii)). As a result, candidates for rows of trees are extracted. Based on the NDVI image, sealed and non-sealed areas in between the trees can be distinguished (hatched areas in Fig. 4 (iii)). The resultant elongated tree/bush regions were selected according to their length. This simple model led to satisfying results in this study.

Roads are modelled as lines. It should be noted that roads can have a higher or lower reflectance than their surroundings. Geometry is explicitly introduced into the model of the road by the assumption that roads are composed of long, straight and horizontal segments. Roads are also described in terms of topology; the road segments form a network, in which all segments are linked to each other. The extraction strategy is derived from the model and is composed of several steps. After line extraction according to Steger (1998), post-processing of the lines is performed with three tasks in mind: (1) increase the probability that lines either completely correspond to roads or represent linear structures which are not roads; (2) fuse lines; and (3) prepare lines for the generation of junctions. A weighted graph is then constructed from the lines and the gaps between them. The weights are derived from radiometric and geometric criteria. From the weighted graph, the road network is extracted by selecting best paths between various pairs of points which are assumed to lie on the road network with a high degree of probability (for a more detailed description of the road extraction algorithm see Wiedemann, 1999).

In the next step, the two object models are combined by applying local context knowledge. A row of trees may either be a specialization of a road segment or only a part of the open landscape. The decision depends on the local context, as follows.

(1) If a row of trees is detected, which lies near and parallel to a road segment, the row of trees is considered part of the open landscape, and no further refinement is necessary; both objects are valid.
(2) If a row of trees is detected and no road segment is nearby, the percentage of sealed pixels in the hatched region in Fig. 4(iii) is determined. If this percentage is high, the row of trees is considered as a candidate for a road segment; in other words, it can be assumed that the trees occlude the road.

FIG. 5. Small portion of the test image which includes a row of trees.

Otherwise no decision can be taken, since the row of trees can either completely occlude the road or can stand back from the road.

Further analysis of the local surroundings can yield more insight: if the row of trees indeed occludes parts of the road, unconnected road segments at both ends of the row of trees should be present. This last step in the analysis, however, has not yet been implemented in this approach. Instead, every instance of a row of trees is introduced as a possible road segment into the weighted graph from which the road network is extracted.

Results

The approach described has been applied to the test image. The first procedure consisted of the extraction of global context and the rows of trees. Fig. 5 shows part of the test image, containing a forest area at the top and open landscape, including a row of trees and an elongated agglomeration of trees, below. The extraction results

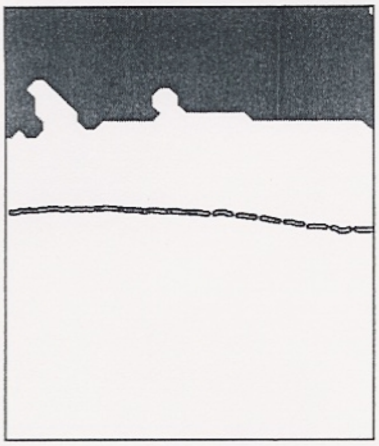

FIG. 6. Extraction results for Fig. 5, showing forest (dark grey), open landscape (white) and a row of trees.

FIG. 7. Subset of the test image, $2 \times 2{\cdot}5$ km^2 in extent.

are depicted in Fig. 6. The global context areas *forest* (dark grey) and *open landscape* (white) were distinguished correctly; the two small intrusions of open ground in the forest are clearings which, at closer inspection, are also visible in the test image. The agglomeration of trees was labelled as open landscape because it is not sufficiently large nor compact enough to be classified as forest. The row of trees was extracted successfully.

Another subset of the test image is shown in Fig. 7. The results of global context extraction followed by road extraction (without an explicit modelling of trees) are presented in Fig. 8. Again, forest is depicted in dark grey, settlement in light grey, open landscape in white, and the grey lines represent the roads. The global context areas were detected correctly, which is also true for the majority of the roads. However, some roads are missing, because the road extraction algorithm was tuned to deliver conservative results. Based on the assumption that it is easier to manually complete an automatically extracted road network than to have to delete parts of the network and to add additional parts, comparatively strict thresholds were used for acceptance of an extracted structure as a road. It can also be seen in Fig. 8 that the two roads occluded by trees, approximately in the centre of Fig. 7, were not extracted. Finally, in Fig. 9, the results of a combined approach are presented. It is possible to see that the roads missing in Fig. 8, due to the occlusions, have now been added to the road network.

The results achieved are promising and demonstrate the advantages of combining the GIS data model with the object model for image analysis, as well as the benefits of combined extraction of more than one object class using a hierarchical strategy. The explicit representation of global context knowledge can be regarded as a framework for the interpretation of an entire scene. The global context for the extraction of objects not treated in the authors' work would also be given, for example the context *settlement* for building extraction. A problem with the combined model, as it stands today, is the implicit modelling of disturbances in the approach used for the extraction of roads. Trees were modelled twice: as a road disturbance and also as part of a row of trees. Further work is needed on the refinement of the object

FIG. 8. Global context areas and roads extracted without modelling trees, showing forest (dark grey), settlement (light grey), open landscape (white) and roads (grey lines).

FIG. 9. Extraction result for a combined approach; additional roads are plotted in black.

models in order to avoid this redundant modelling and to improve completeness. As mentioned, the algorithms were tuned to deliver conservative results. In a comparison with a manual interpretation of the test image, it was found that while the extracted roads and forest areas were correct in about 95 per cent of the cases, only 70 per cent of existing objects were actually extracted. Another aspect that requires attention in the future concerns the integration of watercourses, which look quite similar to a road with bordering trees.

INTERPRETATION AND MONITORING OF MOORLAND

The second example is concerned with the interpretation and monitoring of moorland. This application is important because moorland is an environmentally sensitive habitat and it is also of interest for its industrial value, notably peat extraction. The authors' approach is based on an automatic knowledge-based and rule-based multitemporal interpretation of moorland from aerial CIR images. Up to now, the usual image analysis methods for an interpretation of such areas were data-driven multispectral classifications. However, standard multispectral classification does not use prior knowledge about the area to be interpreted and, therefore, the results are often unsatisfactory. Such prior knowledge includes, for example, the fact that peat extraction is mainly performed by using harvesting machines. These machines leave visible tracks on the ground, which can be recognized as lines in the corresponding images. The use of prior knowledge has the potential to improve interpretation because it reduces the search space by additional constraints.

The interpretation presented is divided into two parts. The first part consists of a monotemporal interpretation and only uses information regarding geometry, radiometry and texture of the land use classes. The second part is an extension to a multitemporal interpretation. The necessary temporal information for this part is formulated in a diagram which describes the most probable state transitions. In the monitoring process, the state transition diagram is used to predict possible land use changes. This procedure leads to a reduction of the search space and improves monitoring.

The system used for the approach presented is the knowledge based system AIDA (Liedtke *et al.*, 1997, Tönjes, 1998), which was developed in order to achieve automatic interpretation of remote sensing images. The system is based on semantic nets for knowledge representation and strictly separates the control of the image analysis process from the semantics of the scene.

Prior Knowledge about Moorland

Some background information about moorland is required in order to understand its peculiarities (see Göttlich, 1990; Eigner and Schmatzler, 1991; Redslob, 1999 for further details). This prior knowledge is subsequently implemented in the knowledge base for the monotemporal and the multitemporal interpretation.

Originally, moors were upland areas. In Germany, this environment has almost vanished. Today, agricultural areas, forests and regions of regeneration or degeneration replace the former upland moors. The most important industrial use of moorland is peat extraction. In order to make peat extraction possible in a moor, the ground first has to be drained and, therefore, ditches need to be created. Thus, the water level goes down and the area begins to degenerate, with consequent vegetation changes. During the state of degeneration, the vegetation is inhomogeneous and irregular. Peat extraction is then possible and harvesting machines are normally used for this task. These machines leave two straight tracks on the ground, which can be recognized as parallel lines in aerial images. After peat working has finished, regeneration of the moorland can begin. In many cases, people simply stop working the land and leave it to regenerate, which eventually results in increased vegetation. Hence, at this regeneration stage of land use, vegetation can again be found on these moorland areas, especially birch trees due to the drying out of the ground. In many cases, vestiges of tracks from the harvesting machines can still be found. In order to encourage regeneration of former upland moor of a similar nature to the original vegetation, supporting measures are sometimes carried out; for example, ditches are filled in and trees are removed in order to raise the water level. If the water level does rise further, trees die and a homogeneous vegetation cover, without trees, appears.

Input Data

The test area used in this study consists of a moorland area to the north-west of Hanover, near Steinhude in Lower Saxony. Aerial images with a ground resolution of 0·5 m have been selected. The main input sources are CIR images that were acquired in summer 1989, but the results were also tested with greyscale images from 1975 and 1988. The reason for this choice is that, although colour images contain more information, most aerial images now available are greyscale images. Also, for the multitemporal approach, greyscale aerial images from different epochs were needed.

The second input source consists of a segment image, which contains the geometric outline of the different segments of the test area which are to be interpreted. The segment image can be based on biotope mapping, which is performed for many moorland areas in Germany by means of manual interpretation of CIR imagery, combined with field checks.

Monotemporal Interpretation

This section describes the monotemporal interpretation of moorland from aerial images. As mentioned before, the system used for knowledge representation is based on semantic nets. Therefore, prior knowledge about the relevant area is formulated according to the demands of such a net. Fig. 10 shows a simplified version of a semantic net.

Moorland is subdivided into four classes: *forest*, *agriculturally used area*, *area of re-/degeneration* and *area of peat extraction*. The classes *area of degeneration* and *area of regeneration* are combined, because their distinction in aerial images from a single epoch is very difficult. As shown in Fig. 10, two layers of abstraction are distinguished: a *scene layer* and an *aerial image layer*. In the scene layer, the different classes are described with their obligatory parts. For instance, the area of peat extraction is characterized by *harvester tracks* and *low vegetation density*. The *area of re-/degeneration* is also characterized by *harvester tracks*, and additionally by *medium vegetation density*. The nodes in the aerial image layer describe the appearance of the scene layer nodes in CIR images. Both colour and texture information are given, so the concept net is suitable for both colour and greyscale images.

At the foot of Fig. 10, *segment analysis operators* are noted. Every node at the bottom of the aerial image layer has access to a special operator. Segment analysis operators serve to verify the meaning of the node for a particular segment by means of image processing. Thus they analyse a given segment and estimate whether a given hypothesis for a node is correct or not. In this way, the operators transform the explicitly formulated hypotheses into image processing operations. Only at this level does the interpretation have direct access to the raster images. As an example, the operator connected to the node *dismembered structure* analyses the structure of the edges in a particular segment. Short and curved lines lead to a better assessment than long and straight lines.

During the interpretation process, the nodes of the semantic net are instantiated with the concept net discussed as prior knowledge. Instantiation starts with a predefined seed node and then proceeds in a particular order along the relationships postulated in the concept net, until no more rules can be applied and instantiation stops. In order to show the instantiation process in the case presented, it is described as follows. The procedure starts with the creation of a hypothesis of the concept

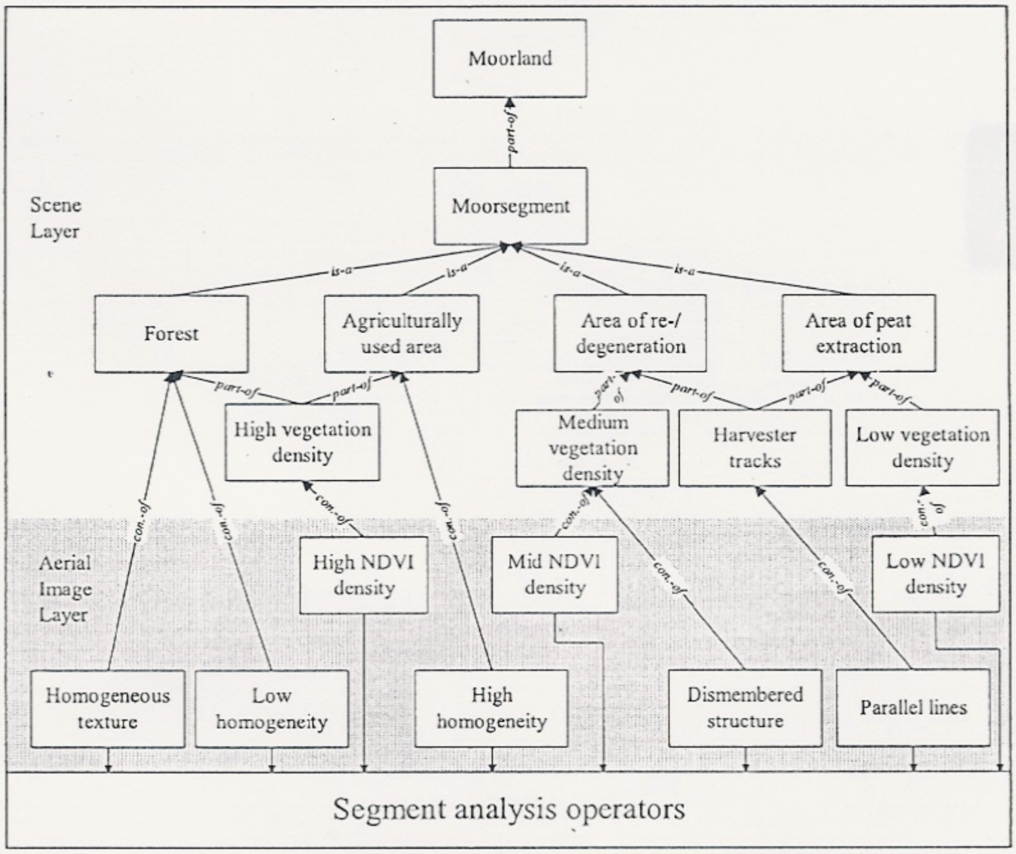

FIG. 10. Semantic net for monotemporal interpretation of moorland.

moorsegment. At this point, one segment is taken from the segment image. The interpretation for this segment is performed next and, as shown in Fig. 10, four different interpretation possibilities for the segment exist. These possibilities exclude each other and therefore compete with each other. The first class to be verified is *area of peat extraction*, so a hypothesis *area of peat extraction* is created. Two obligatory parts of this node have to be present: *harvester tracks* and *low vegetation density*. This pathway leads to the top–down instantiation of the concept *harvester tracks* along the *part-of* relation. In the aerial image layer, *harvester tracks* translates into *parallel lines*, which leads to the creation of a hypothesis *parallel lines*. Now the bottom layer has been reached and this hypothesis has to be verified. The node calls up a special *segment analysis operator*. The operator examines the aerial image within the given segment and responds as to whether parallel lines were found or not. If the result is positive, the operator returns a certainty value for the node, which describes the quality of the result, and the instance node *parallel lines* changes its status from *hypothesis* to *complete instance*. This progression leads to a complete bottom–up instantiation of the node *harvester tracks*. In a similar manner, the second obligatory part of the hypothesis *area of peat extraction*, namely *low vegetation density*, is verified and for the second verification a certainty value is also determined. Now all the obligatory parts of *area of peat extraction* are present and the node is completely instantiated. A certainty value for this node is computed from the values of the nodes below it. The result is a possible interpretation of the *moorsegment*, with

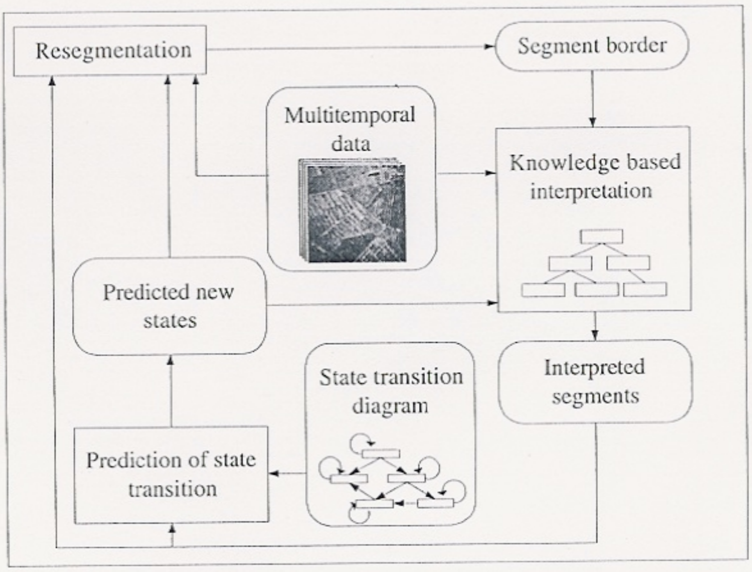

FIG. 11. Concept for multitemporal interpretation of moorland.

a certainty value. If the latter lies below a predefined threshold, the other competing interpretations are checked in the same way.

Multitemporal Interpretation

In this section, the extension of the system to multitemporal interpretation is described. This further operation is necessary for the application described because, besides the assignment of classes (also called *states* in applications involving temporal change) to areas in moorland, the monitoring of change is very important. The multitemporal interpretation begins with an initial interpretation for the aerial images taken at some epoch t. The next epochs then have to be interpreted, based on the results of the previous interpretation.

Fig. 11 shows an overview of the multitemporal system. Beginning with *knowledge based interpretation*, an initial interpretation of the segments is performed. The results are *interpreted segments* of moorland. These segments are then input for a *prediction of state transition*. This prediction uses prior information concerning possible changes and the possibilities are represented in a *state transition diagram*. The outputs of the prediction are *predicted new states* for every segment. The borders of the segments may change between the interpretation intervals. Therefore, for the multitemporal approach, a module is included to perform segment splitting by segmentation. The approach uses the information of the *predicted new states* and is described in detail in Pakzad *et al.* (1999). The results of this step are updated segment borders, which are integrated into the *knowledge based interpretation* for the new epoch, as are the predicted new states and the multitemporal data.

The temporal part of the prior knowledge is implemented in the *state transition diagram*, as illustrated in Fig. 12. It describes the most probable state transitions. Although many more state transitions are theoretically possible, they are limited by legal and natural constraints, and it is possible to use these restrictions in order to improve the interpretation. In contrast to the concept net shown in Fig. 10, the state

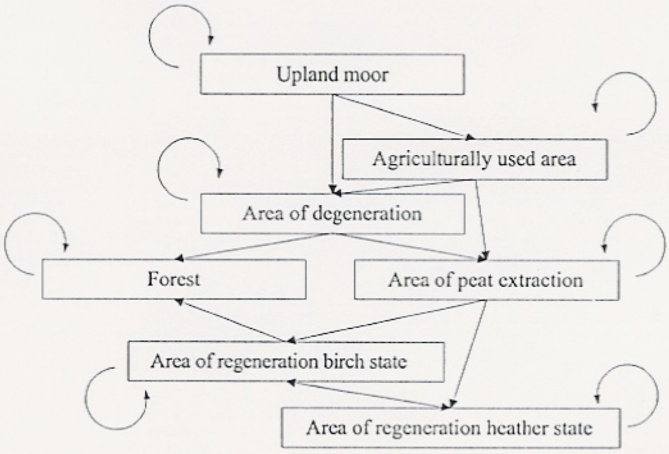

FIG. 12. State transition diagram.

transition diagram contains seven different states. The first state, *upland moor*, is included only to complete the diagram. The *area of degeneration* and *area of regeneration* are now separated. In addition, the *area of regeneration* is divided into two substates: *birch state* and *heather state*. As mentioned, the distinction between the *area of degeneration* and *area of regeneration* in aerial images taken at one epoch is very difficult. In a multitemporal interpretation, however, the development of the different segments can also be used. For example, given an *area of peat extraction*, where the system knows that this segment has passed the *area of degeneration* state, if in a new epoch a segment analysis operator finds vegetation, the only possible states are *area of regeneration* and *forest*. Prior knowledge also allows the distinction to be made between the two regeneration states. If no more *parallel lines* can be found on an area which was previously used for peat extraction, and if the system finds homogeneous areas, the system expects that no more ditches exist to drain the ground, and the requirement for the *area of regeneration: heather state* is correct. Otherwise, an *area of regeneration: birch state* is postulated.

Every link in the state transition diagram has a priority, which models the probability of the state transitions over time. The probabilities also depend on the time difference between two epochs and they affect the order in which the different state transition hypotheses are verified. As shown in Fig. 12, every state has a transition link back to itself, which is the link with the highest probability. Consequently, for every new epoch this is the first transition concept to be verified. The semantic net used for the multitemporal interpretation is a refinement of the net depicted in Fig. 10 and is shown in Fig. 13. In order to model the knowledge that a certain property must be absent (for example the parallel lines in areas previously labelled as peat extraction), an additional relationship, the negation *not* is introduced in the semantic net.

Results

In Fig. 14, parts of the test area as depicted in the CIR image are shown, together with the results achieved for monotemporal interpretation. The biotope mapping used for segmentation has been carried out, based on the same imagery. It was found that all 33 segments were interpreted in the same way as an experienced human operator would interpret them, using only the CIR image. The selection of greyscale aerial

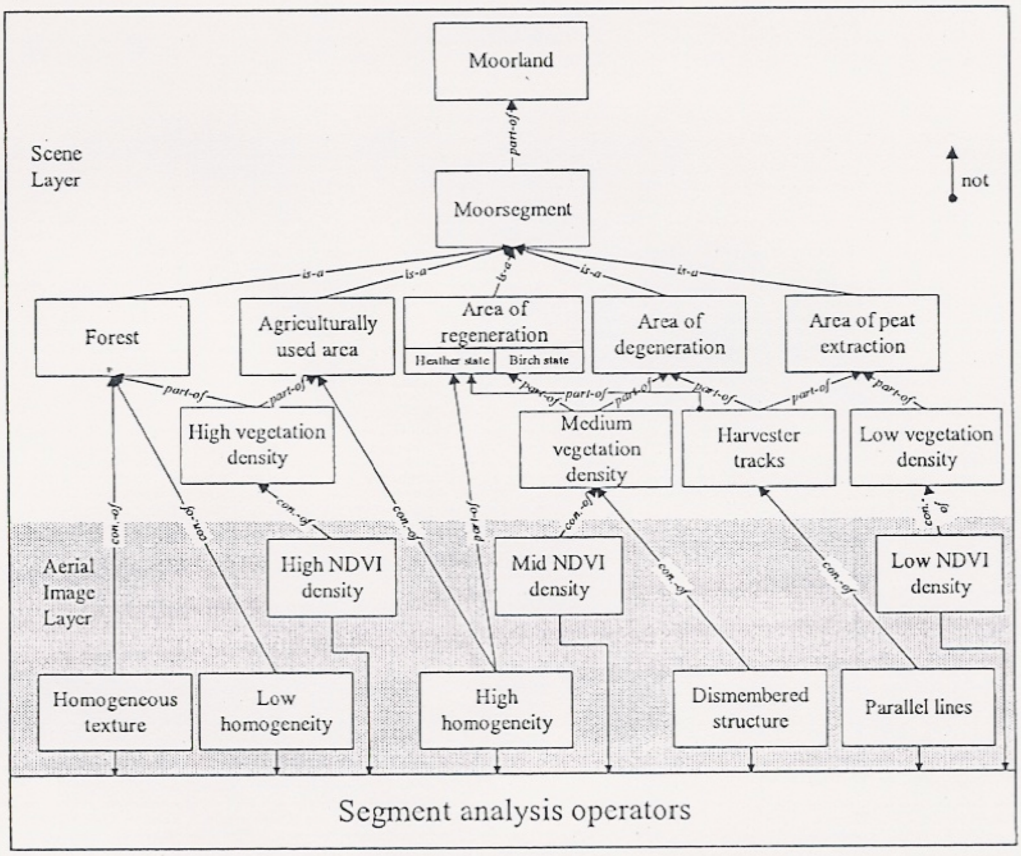

FIG. 13. Extended semantic net for multitemporal interpretation.

images, instead of CIR images, led to similar results (Pakzad *et al.*, 1999). Thus, while colour information in general contains additional information, texture information is sufficient for the interpretation for most of the regions in the example given.

For a small part of the test area, a multitemporal interpretation was performed. As input data, two greyscale aerial images taken in 1975 and 1988 were available; the initial segmentation was carried out manually and led to five segments. For every segment, the system determined the state transitions. Fig. 15 shows the results obtained. For segment 1, a transition from *area of peat extraction* to *area of regeneration: heather state* was determined, using knowledge about the previous land use of the segment and about the missing parallel lines. It should also be noted that, based on the multitemporal imagery, a distinction between *area of regeneration* and *agriculturally used area* is possible for segment 1, although both states have identical descriptions in the semantic net, when only black and white images are available and NDVI information cannot be used (Fig. 13). For segment 2, the land use state changed from *area of peat extraction* to *forest*, although there is no direct state transition between the two states in the state transition diagram. Due to the rather lengthy time interval of 13 years between 1975 and 1988, the state *area of regeneration* was probably not observed. Using the knowledge about the mean transition times, the system generated the hypothesis for *forest*, which was then successfully verified. For the other three segments no transition was detected.

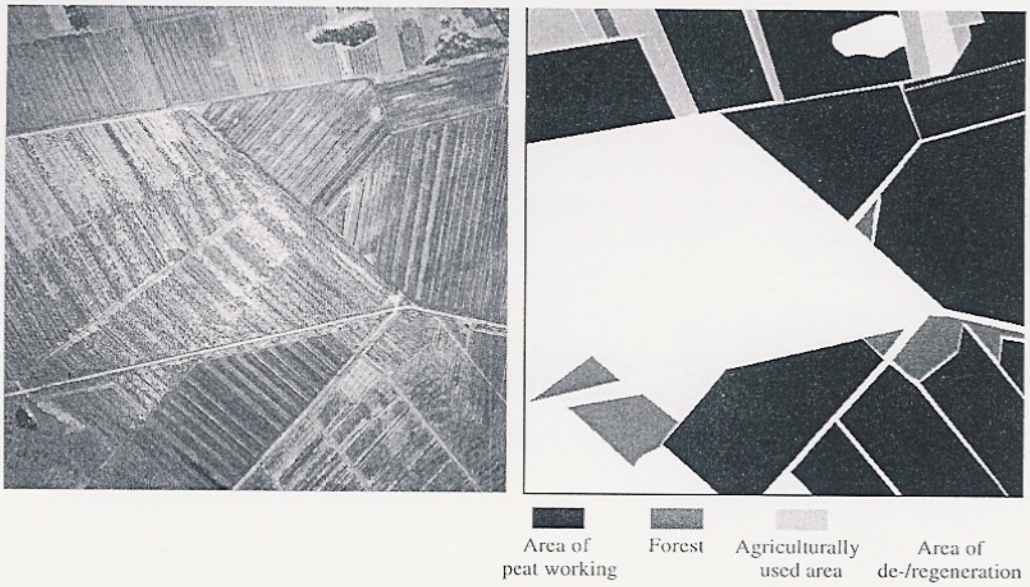

FIG. 14. Aerial images of the test area and results of monotemporal interpretation.

It could thus be shown that multitemporal interpretation enables distinction to be made between more land use classes as compared to monotemporal interpretation. This deduction applies, for example, to the distinction between the land use classes *area of regeneration* and *area of degeneration*. In order to distinguish between these two states, temporal knowledge, in other words whether peat extraction has already been carried out or not, is necessary. The exploitation of temporal knowledge also leads to a more robust interpretation of land use classes, for example the distinction between *agriculturally used area* and *area of regeneration* in black and white images. The use of temporal knowledge can, therefore, partly replace a need for colour information.

The results presented constitute a proof of concept for monotemporal and for multitemporal interpretation of moorland. In the future, research will continue in various directions. Currently, the image processing operators depend on parameters which have to be adapted manually to the images used. One of the future objectives is to automate this parameter tuning. Work is also being carried out on a refined modelling of moorland with more object classes and more attributes, on substituting the biotope mapping by a segmentation module working only on the imagery, and on a refinement of the currently rather crude uncertainty management system.

CONCLUSIONS

Research into image analysis concerned with topographic applications has been described. Key aspects of the work include the integration of the models for object extraction and the GIS data models, the hierarchical reduction of the image complexity, the combined extraction of different objects providing mutual support for each other, and multitemporal image interpretation and monitoring using state transition diagrams. In the authors' experience, a hierarchical processing strategy making use of global and local context, and combining statistical image analysis at a coarse resolution level with model-based approaches at a fine level is particularly important. They are convinced that the approaches described are not only valid for aerial CIR

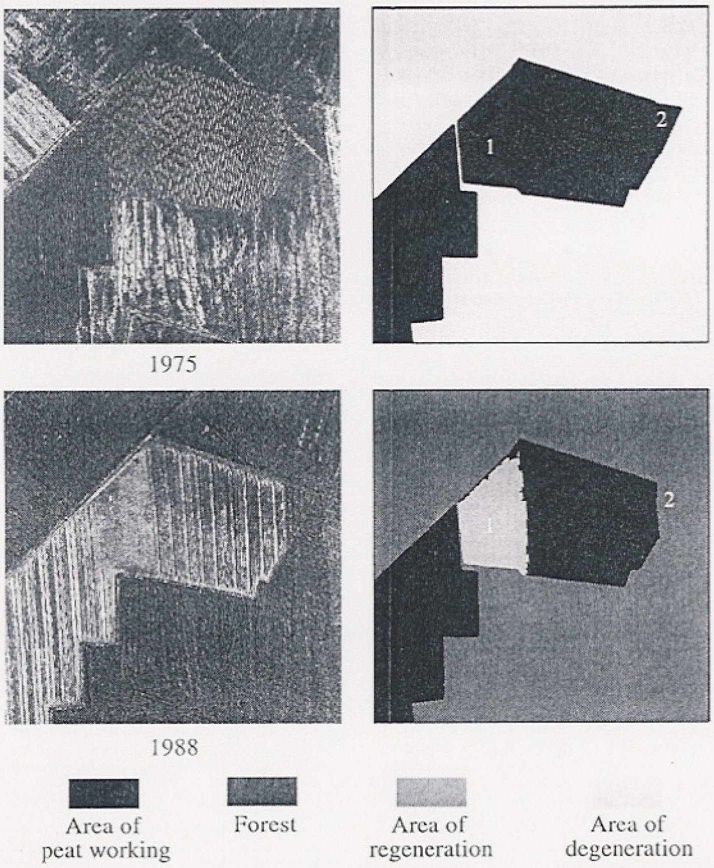

FIG. 15. Results of multitemporal interpretation. Transition was detected in the segments 1 and 2 marked, but not in the other three segments that were tested.

images, but also for satellite imagery with similar geometric and radiometric characteristics, such as that acquired by IKONOS or by other forthcoming high resolution missions.

ACKNOWLEDGEMENTS

The authors would like to express their thanks to C. Wiedemann, Technical University, Munich for making the results of his road extraction algorithm available. Portions of this work were developed within the IST Programme CROSSES financed by the European Commission under the project number IST-1999–10510. Support was also obtained from the Deutsche Forschungsgemeinschaft under grant Gr 1152/3.

REFERENCES

BAUMGARTNER, A., ECKSTEIN, W., MAYER, H., HEIPKE, C. and EBNER, H., 1997. Context supported road extraction. In *Automatic extraction of man-made objects from aerial and space images (II)* (Eds. A. Gruen, E. P. Baltsavias and O. Henricsson). Birkhäuser, Basel. 393 pages: 299–308.

BORGEFORS, G., BRANDTBERG, T. and WALTER, F., 1999. Forest parameter extraction from airborne sensors. *International Archives of Photogrammetry and Remote Sensing*, 32(3–2W5): 151–158.

EBNER, H., ECKSTEIN, W., HEIPKE, C. and MAYER, H. (Eds.), 1999. Proceedings, ISPRS Conference on Automatic extraction of GIS objects from digital imagery. *Ibid.*, 32(3–2W5). 365 pages.

ENGLISCH, A. and HEIPKE, C., 1998. Erfassung und Aktualisierung topographischer Geo-Daten mit Hilfe analoger und digitaler Luftbilder. *Photogrammetrie Fernerkundung Geoinformation*, 3/1998: 133–149.

EIGNER, J. and SCHMATZLER, E., 1991. *Handbuch des Hochmoorschutzes—Bedeutung, Pflege, Entwicklung.* Kilda-Verlag, Greven. 158 pages.

FISCHER, A., KOLBE, T., LANG, F., CREMERS, A. B., FÖRSTNER, W., PLÜMER, L. and STEINHAGE, V., 1998. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Computer Vision and Image Understanding*, 72(2): 185–203.

FÖRSTNER, W., LIEDTKE, C.-E. and BÜCKNER, J. (Eds.), 1999. *Proceedings, Workshop on Semantic modelling for the acquisition of topographic information from images and maps (SMATI'99).* 169 pages.

GÖTTLICH, K., 1990. *Moor- und Torfkunde.* Schweizerbart Verlag, Stuttgart. 529 pages.

GOUGEON, F. A., 1995. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing*, 21(3): 274–284.

HEIPKE, C. and STRAUB, B.-M., 1999. Relations between multiscale imagery and GIS aggregation levels for the automatic extraction of vegetation areas. *ISPRS Joint Workshop on Sensors and mapping from space 1999.* Institute for Photogrammetry and Engineering Surveys, University of Hanover. 6 pages. Available on CD-ROM.

LIEDTKE, C.-E., BÜCKNER, J., GRAU, O., GROWE, S. and TÖNJES, R., 1997. AIDA: A system for the knowledge based interpretation of remote sensing data. *3rd International Airborne Remote Sensing Conference and Exhibition,* II: 313–320.

LINDEBERG, T., 1994. *Scale-space theory in computer vision.* Kluwer Academic Publishers, Boston, USA. 423 pages.

MAYER, H., 1998. Automatische Objektextraktion aus digitalen Luftbildern. *Deutsche Geodätische Kommission*, Series C, 494. 132 pages.

NIEMANN, H., SAGERER, G. F., SCHRÖDER, S. and KUMMERT, F., 1990. ERNEST: a semantic network system for pattern understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9): 883–905.

PAKZAD, K., BÜCKNER, J. and GROWE, S., 1999. Knowledge based moorland interpretation using a hybrid system for image analysis. *International Archives of Photogrammetry and Remote Sensing*, 32(3–2W5): 159–165.

PINZ, A., 1989. Final results of the vision expert system VES: finding trees in aerial photographs. In *Wissensbasierte Mustererkennung* (Ed. A. Pinz). OCG-Schriftenreihe, Oldenbourg. 234 pages: 90–111.

RAPP, F., 1995. Modell und Realität. *Zeitschrift für Photogrammetrie und Fernerkundung*, 63(6): 220–223.

REDSLOB, M., 1999. *Radarfernerkundung in niedersächsischen Hochmooren.* Dissertation, Institut für Landschaftspflege und Naturschutz, Universität Hannover. 278 pages.

RICHARDS, J. A. and JIA , X., 1999. *Remote sensing digital image analysis: an introduction.* Third edition. Springer Verlag, Berlin. 363 pages.

ROSENFELD, A., 1982. Computer image analysis: an emerging technology in the service of society. *Computer Science Technical Reports TR-1177, MCS-79-23422*, University of Maryland. 10 pages.

SAGERER, G. and NIEMANN, H., 1997. *Semantic networks for understanding scenes—advances in computer vision and machine intelligence.* Plenum Publishing Corporation, New York. 512 pages.

SUETENS, P., FUA, P. and HANSON, A. J., 1992. Computational strategies for object recognition. *ACM Computing Surveys*, 24(1): 5–60.

STEGER, C., 1998. An unbiased detector of curvilinear structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(2): 113–125.

TÖNJES, R., 1998. *Wissensbasierte Interpretation und 3D-Rekonstruktion von Landschaftsszenen aus Luftbildern.* Dissertation, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover. 117 pages.

WIEDEMANN, C., 1999. Completion of automatically extracted road networks based on the function of roads. *International Archives of Photogrammetry and Remote Sensing*, 32(3–2W5): 209–215.

Résumé

L'interprétation automatique de l'imagerie aérienne et spatiale par l'analyse d'images est actuellement l'un des thèmes majeurs de recherche en photogrammétrie et dans les disciplines connexes. Le but principal est l'extraction automatique des objets topographiques visibles, dimensionnés, de cette imagerie. Ces objets, une fois saisis, constituent une forme d'entrée possible dans la création de bases de données géographiques. On examine dans cet article divers aspects de l'analyse d'images et l'on fournit un cadre pour l'interprétation des scènes, en se basant sur l'intégration de