TOWARDS THE AUTOMATIC INTERPRETATION OF IMAGES FOR GIS UPDATE

Bernd – M. Straub, Christian Wiedemann*, Christian Heipke
Institute for Photogrammetry and Engineering Surveys, University of Hanover,
Nienburger Strasse. 1, 30167 Hanover, Germany
straub@ipi.uni-hannover.de, heipke@ipi.uni-hannover.de
*Chair for Photogrammetry and Remote Sensing, Technische Universität München,
Arcisstrasse 21, 80333 Munich, Germany
wied@photo.verm.tu-muenchen.de

IC Working Group IV/III.2 Integration of Image Analysis and GIS

KEY WORDS: GIS, Image Understanding, Integration, Model-based Processing, Semantics

ABSTRACT:

The automatic interpretation of aerial and satellite imagery is one of the main research tasks in photogrammetry and related disciplines. The primary goal is the automatic extraction of visible, spatial, topographic objects from imagery. The extracted objects are one possible input for updating or creating geographic data bases. In order to simplify the task, automatic image interpretation was subdivided into more specialised problems like the extraction of buildings or roads and natural objects like vegetation or water. A lot of very promising approaches were presented in the past, but at the same time the interpretation of the whole scene was not tackled in depth. In this paper we discuss different aspects of this issue and provide a frame for scene interpretation which is based on a GIS data model. Results of the proposed approach dealing with roads and trees and their interrelations are presented.

1 INTRODUCTION

Image understanding in the field of topographic mapping is a main research task in photogrammetry and computer vision, see (GÜLCH & MAYER, 1999; EBNER et al., 1999; FÖRSTNER et al., 1999) for an overview. Most of the approaches were designed for the extraction of one object class only, e.g. roads, houses, or vegetation. These individual subtasks can be carried out successfully as long as it is ensured that the image actually contains objects of the respective class which are well separated from other - not modelled - objects. However, the separation into subtasks leads to restrictions in more complex imagery, in which different object classes appear close to and disturb each other. In order to – partly – overcome these restrictions, two approaches – for the extraction of roads and vegetation, respectively – were integrated into one model.

The key idea of this work is that a landscape can be subdivided into a few abstract superclasses. Assuming that the observed part of the landscape is large enough to include instances of all these superclasses, a strategy for their determination is proposed. A specific object extraction algorithm can then be applied within those superclass areas, in which they are known to work successfully, and local context can support the extraction of different objects. The described approach can be applied to capturing and updating GIS data bases for topographic mapping.

In the following we begin with a short description of the used data sources. Afterwards, the general strategy is explained by means of the used models for the extraction of roads and trees as well as their integration into one common model. Subsequently, first results of the combined approach are presented.

2 DATA SOURCES, DATA MODEL, AND APPLICATION BACKGROUND

In the following we propose a frame for the automatic interpretation of multispectral aerial or satellite imagery with a ground sampling distance (GSD) of about 1[m]. The images, which were used in this study, have been acquired in summer 1997 using the DPA sensor (Digital Photogrammetric Assembly, see FRITSCH, 1997 for details). They have a GSD of approximately 0.8[m]. Reduced resolution images were computed in morphological and linear scale space with 1.5[m] and 12[m] GSD, respectively. The DPA camera operates in four spectral bands. Three bands (red, green, blue) are in the visible range of the spectrum and the fourth one lies in the near infrared. Other imagery, e.g. from satellite based sensors with similar characteristics, like IKONOS (SPACE IMAGING, 2000), in combination with image merging algorithms (e.g. SCHIEWE, 1999; WALD et al., 1997) could be used as well.

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

The size of the investigated scene is about 35[km²]. It is thus large enough to contain one or more villages, forest, and agricultural areas as well as parts of the road network. The variety of topographic objects is a prerequisite for our approach, which in Germany and in Central Europe is often fulfilled within a few square kilometres. The detailed interpretation of settlement areas including a reconstruction of buildings, roads, and trees is out of the scope of this paper. Nevertheless, these tasks can be integrated into the strategy, but then, inside of settlement areas, a higher image resolution of up to 0.1[m] and preferably also height data are necessary.

3 REDUCTION OF COMPLEXITY – A POSSIBLE FRAME FOR AUTOMATIC IMAGE INTERPRETATION

In this chapter, the relation between the information content of imagery in fine and coarse scale and aggregation levels in a *GIS data model* are discussed. In general the interpretation of images is based on a model of the real world, the *model for object extraction*. Following the proposal of (SUETENS et al., 1992, also used in MAYER, 1998 and MAYER, 1999) the methods of image analysis can be mapped into a two-dimensional solution space (see figure 1), one axis describes the *suitability for complex models* and the other axis the *suitability for complex 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, indicated by "?" in figure 1. 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 and by global context knowledge, i.e. by restricting the image area to be processed.

We argue that in order to solve a problem with complex models and complex images, 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. For example, a simple model is that of multispectral classification.

(ii) Reduce the image complexity using global context knowledge in the form of the results of (i) to the images of higher resolution. Within a global context area the complex model needed to solve the given task can be applied more successfully than in the complete image.

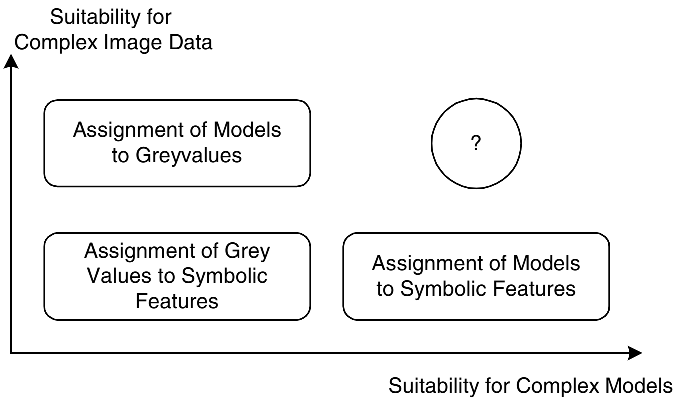

Figure 1: Solution Space of different Methods for Image Analysis

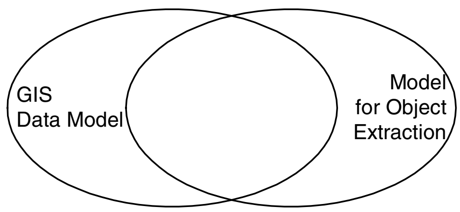

Figure 2: Intersection of Models

The global context defines the frame for the extraction of individual objects, and makes the automatic interpretation of higher resolution imagery more feasible (see also RAPP, 1995). An example for an early implementation of this strategy is the multi scale road extraction approach of (BAUMGARTNER et al., 1997). Road extraction is only performed in those global context areas, where it is known to work reliably. The further analysis then depends on *local context*, i.e. relations between objects.

The GIS data model for the objects is in many ways similar to the model for object extraction, which is not surprising because both models describe the same objects. For the automatic interpretation, the GIS model has to be enhanced with the aim to relate the description of the real world with the imagery. On the other hand the GIS data model contains aspects which are not visible in the image (see figure. 2). The use of the extracted objects as input or update information for GIS data bases becomes much easier if the intersection between these two models is made as large as possible and well defined. To achieve these goals we have defined four superclasses which all lie in the intersection of both models: *settlement*, *water*, *forest* and *open landscape* (HEIPKE & STRAUB, 1999b). In many GIS data models, quite similar classes can be found as the highest level of abstraction in a hierarchical structured data model. Therefore they can be used both in the GIS domain and in the domain of object extraction. These superclasses can be used as *global context* knowledge about the observed subset of the real world. The global context knowledge can be obtained either from external sources such as existing GIS data or from imagery at a reduced GSD, which leads to an abstraction of the image content (LINDEBERG, 1994; LINDEBERG, 1996). In order to be independent of existing GIS data bases we propose an approach based on the imagery alone, using an enhanced multispectral classification, see chapter 4 for details. The reason for having chosen this approach is that the existing GIS data may be outdated and might disturb the interpretation.

In this paper we enhance the strategy of (BAUMGARTNER et al., 1997) by means of *explicit* modelling of local context knowledge which is given by trees occluding a road. In order to be able to use this knowledge, trees are automatically extracted, again using global context. These explanations lead to the following strategy for the interpretation of images (see figure. 3). After a first reduction of the image data complexity by means of a scale space transformation the global context is extracted. According to the global context the image domain is reduced in the fine scale, which also leads to a reduction of the image data complexity. Within each global context class the extraction of objects in the fine scale is performed. With reference to figure 1 this strategy means that the individual approaches should work as near as possible to the origin of the co-ordinate system, given a fixed complexity of the model.

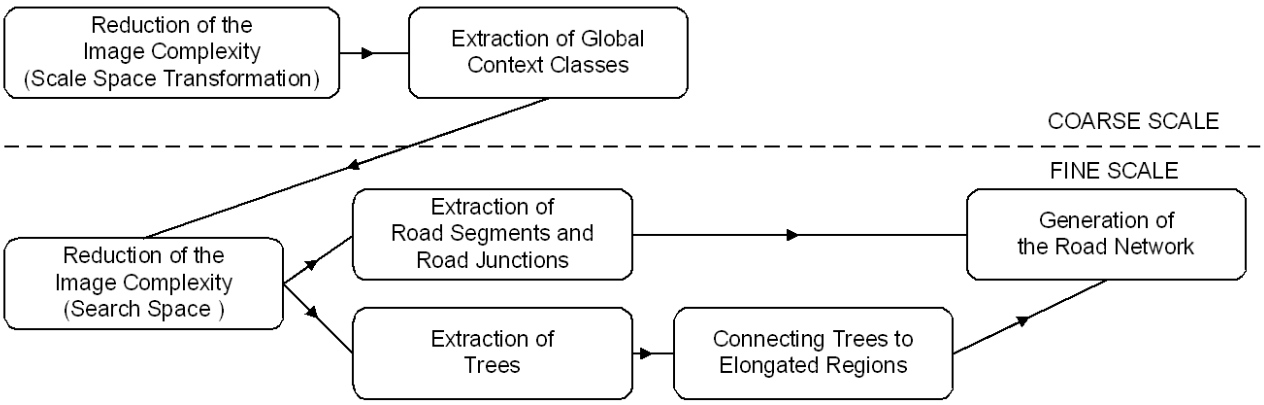

Figure 3: Strategy for the interpretation based on global and local context in different scales

4 GLOBAL CONTEXT – EXTRACTION AND APPLICATION

Global context is described by a simple model in the form of a semantic network, which contains the four superclasses settlement, water, forest, and open landscape (see Figure. 4). This is in contrast to the relatively complex content of high resolution images. To overcome this contradiction, a transformation in morphological scale space is applied to simplify the image content. The scale space transformation can be looked upon as an abstraction of the image content, e.g. single buildings in settlement areas are aggregated to a homogeneous region (STRAUB & HEIPKE, 1999b). The morphological scale space is related to the size of the objects, therefore, it is relatively independent of sensor characteristics and the actual lighting conditions of the scene. In the following the reduced image resolution of 12[m] in morphological scale space is used.

4.1 Extraction of the Global Context Knowledge

Global context knowledge is extracted by an enhanced multispectral classification. The training areas for the multispectral classification are derived automatically as follows. Based on the Normalised Difference Vegetation Index and the local variance settlement and water areas were differentiated (see also, KUNZ & VÖGTLE, 1999). The remaining areas are subdivided into three classes by means of a cluster analysis (ISODATA, see RICHARDS & JIA, 1999), the darkest region contains forest regions and the two lighter regions contain the open landscape. Next, the largest regions in each class were selected and used as training areas for a multispectral classification. Subsequently geometric constraints were applied based on the object description in the GIS data model. The GIS data model defines explicitly the minimum size of the objects and implicitly the compactness, these constraints were used to refine the results of the multispectral classification. The instances of the object classes settlement and forest, initially stemming from the multispectral classification, were selected by means of these geometric constraints. All rejected areas, i.e. roads in open landscape which posses spectral characteristics similar to settlement; and trees and bushes which have been classified as forest, were looked upon as open landscape. In the coarse scale we assume that they are used for agricultural purposes.

4.2 Application of Global Context Knowledge

As described in chapter 3 the global context knowledge is used to reduce the complexity of the image content. The applied road extraction algorithm (WIEDEMANN, 1999) is known to work reliably only in the open landscape. In forests, roads are often not visible due to occlusions by trees. In settlement areas on the one hand, other objects like houses or drive ways have a similar appearance as roads. On the other hand, the appearance of roads is disturbed by many shadows, cast by high objects like trees or houses as well as by occlusions caused by these objects or by cars. Therefore, the reduction of the search space for the extraction of roads to the open landscape only, ensures a more correct extraction result. Referring to figure 1 the complexity of the image data is reduced in order to establish convenient conditions for the road extraction.

We select trees as the vegetation object in this study, the reasons for this selection were: (1) in the context *open landscape* there is an obvious relation between roads and trees, namely "trees may occlude roads" (2) we can demonstrate a further application of global context knowledge, "the colour of trees is given by the colour of forest" and (3) the ATKIS feature class catalogue defines an object *Row of Trees*.

The relation "tree may occlude roads" leads to the demand that trees and roads should be extracted in two independent steps, connected only by the context open landscape. The knowledge of forest area from the global context extraction gives us the possibility to approximate the spectral appearance of trees in the fine scale.

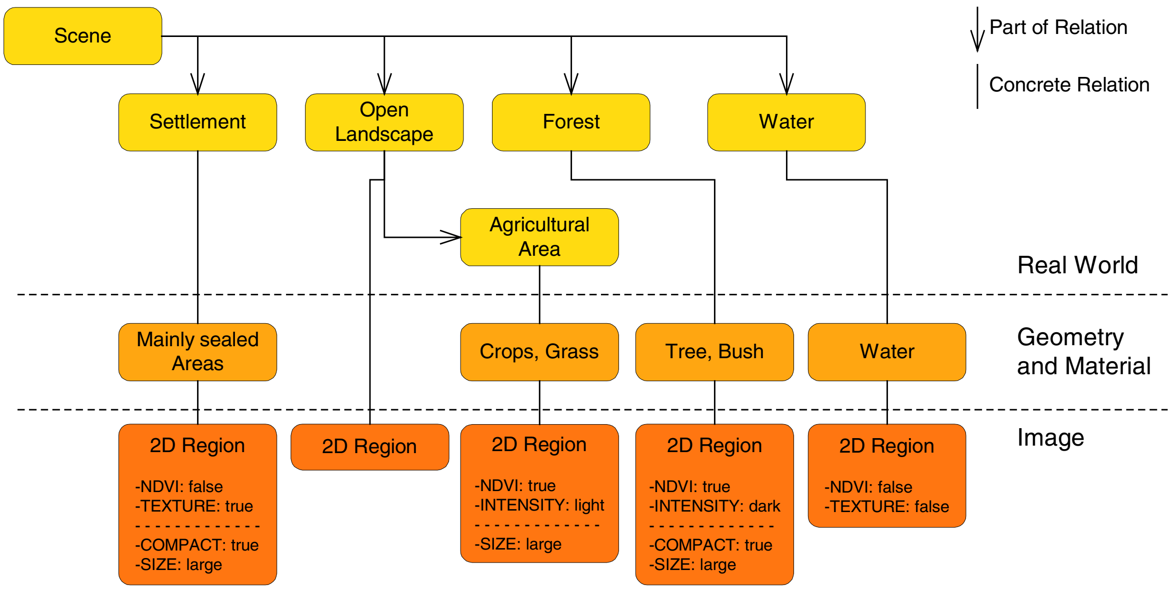

Figure 4: Model of Global Context Classes

5 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. The approach for the extraction of roads is embedded into the global context and it is based on the function of the road network. The model for the extraction of trees is based on the object description in the given GIS data model. The resulting objects *Row of Trees* can be a specialisation of a road segment and/or a part of the open landscape.

5.1 Separate Extraction of Objects

The **road** is modelled as a line. It can have a higher or lower reflectance than the 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 topologically linked to each other. The extraction strategy is derived from the model and is composed of different steps.

After line extraction according to (STEGER, 1998) postprocessing of the lines is performed with three tasks in mind: (1) increase the probability that lines either completely correspond to roads or to linear structures not being roads, (2) fuse lines, (3) prepare lines for the generation of junctions. Then a weighted graph is 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 the selection of best paths between various pairs of points which are assumed to lie on the road network with high probability (WIEDEMANN, 1999).

As described above the model for **trees** is taken from the object description in the GIS data model. In the feature class catalogue we find the following constraints for the feature class 4201 *Row of Tree*. The data capture criteria 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, we use the diameter of a single tree. Single trees are not modelled in the ATKIS feature class catalogue, therefore we have to add a model of a generic individual tree. A simple model is sufficient for our task, more detailed tree models were suggested in the literature, e.g. a rule based approach for the delineation of crowns (GOUGEON, 1995) or the searching of local brightness maxima (PINZ, 1989). Our single tree is a round blob with a diameter of about 10[m] and a characteristic colour. The extraction of instances of the object "Row of Trees" in image space is executed as follows. First blobs with the specified size and the spectral appearance of *forest* areas are searched for inside the *open landscape* region (see Part (i) in Figure 6). These regions were dilated in order to span the distance between single trees (Part (ii) in figure 6). The resulting elongated tree / bush regions were selected according to of their length. This simple model led to satisfying results in our study.

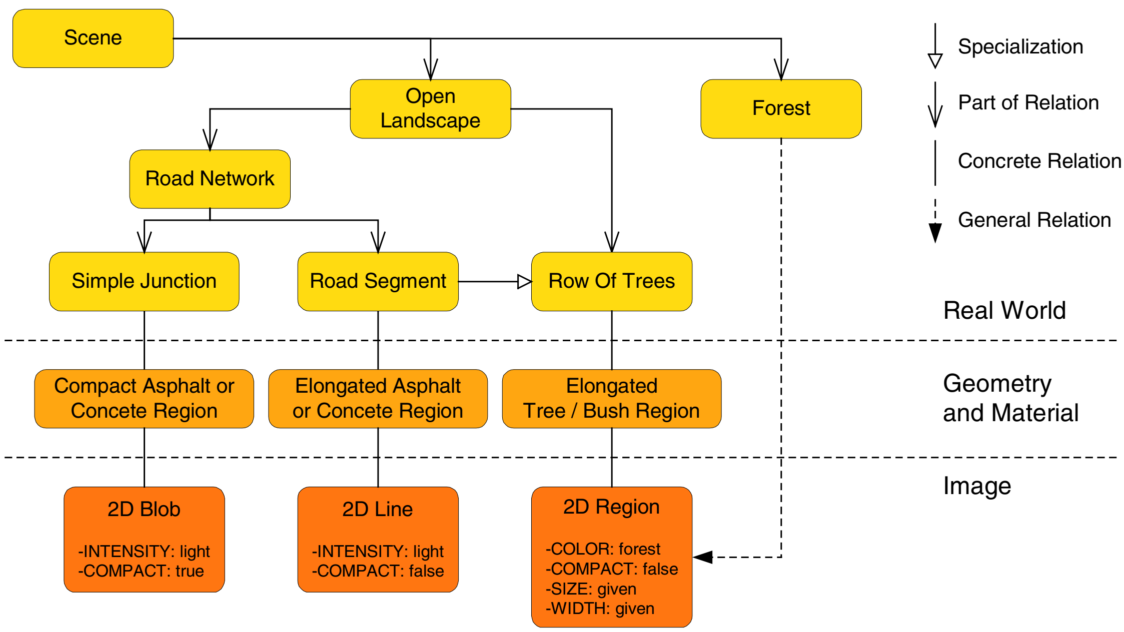

Figure 5: Model for the Extraction of the Road Network and Row of Trees

5.2 Combination of Objects by Applying Local Context

Figure 5 gives an overview of the models integrated into the global context classes and their interrelations based on local context. A *Row of Tree* object may be either an instance of a road segment or it is a part of the open landscape, or both. Possible cases for the decision, depending on *local context*, are: A row of trees was detected, and lies near and parallel to a road segment. In this case no further refinement is necessary, both objects are valid. A Row of Trees was detected and no road segment is nearby. In this case the decision depends on the percentage of sealed pixels in the hatched region in part (iii) of figure 6. In our implementation every instance of a *Row of Tree* object is introduced as a possible road segment into the weighted graph from which the road network is extracted.

A problem in the combined model is the implicit modelling of disturbances in the used approach for the extraction of roads (BAUMGARTNER et al., 1997). Trees were modelled twice – as a disturbance and as a valid object. As a consequence the roads were detected without the explicit knowledge of the trees, this kind of redundancy should be avoided.

A subset of the processed imagery is shown in figure 9, figure 10 shows the results of the global context extraction followed by road extraction without an explicit modelling of the trees. Forest is depicted in dark grey (green in the electronic version of this paper), settlement in light grey (pink in the electronic version), open landscape in white, and grey lines represent the roads. In figure 11 the results of the combined approach are presented. One can see that the road occluded by trees in the centre of figure 9 was extracted successfully.

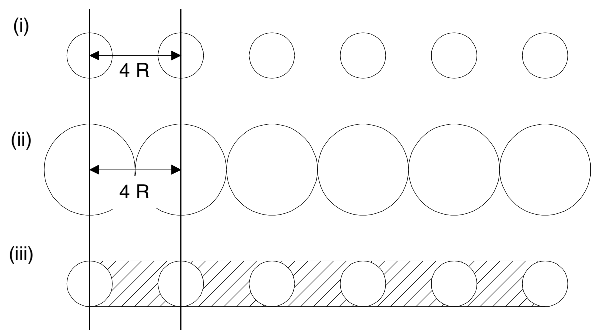

Figure 6: Row of Trees – Parameters and Strategy

Figure 7: A Row of Trees in the image

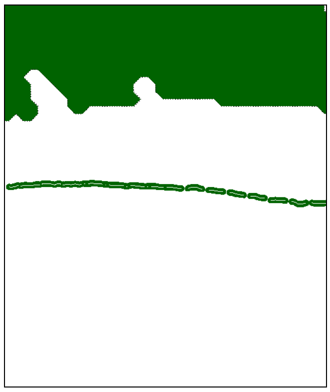

Figure 8: Symbolic features, *Row of Trees, Open Landscape and Forest*

Figure 9: Subset of the processed image, 2*2.5[km²] in size.

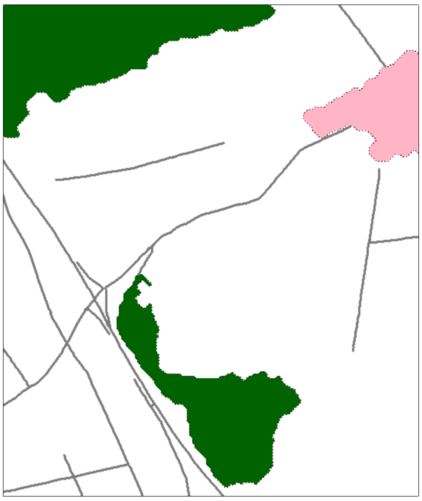

Figure 10: Global context areas and extracted roads without explicit Row of Trees modelling

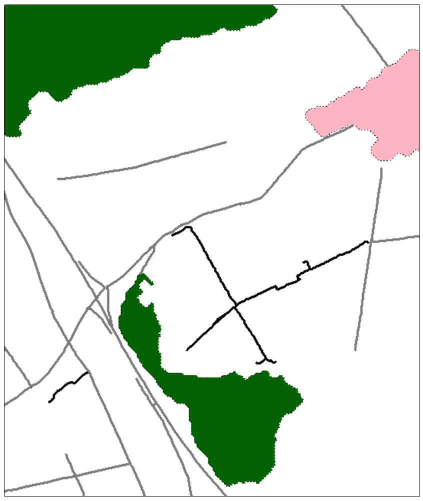

Figure 11: Result of the combined approach, additional roads are plotted in black.

6 EVALUATION OF THE RESULTS

In this section the results of the different steps were compared with a reference interpretation. The reference interpretation was executed by a human operator. The results of the approach were evaluated by means of two global quality measures - *completeness* and *correctness*. The completeness is the percentage of the reference data (REF) which is explained by the automatically extracted data, and the correctness represents the percentage of correctly extracted objects (EXT). The input quantities for these measures, TP (True Positive), FP (False Positive), and FN (False Negative) can be defined by means of entities as follows:

TP: All entities of one object class which are common in both data sets, REF and EXT.
FP: All entities of one object class which are members of the data set EXT but not included in REF.
FN: All entities of one object class which are members of the data set REF but not included in EXT.

It should be noted that the definition of these quantities depends on the different objects. In the case of areas, here the instances of the global context classes, the number of pixels in each category was counted. In the case of roads the comparison is carried out by matching both data sets, EXT and REF, using a so called "buffer method" (see WIEDEMANN et al., 1998 for details).

	$\underset{[0,1]}{Completeness} = \frac{TP}{TP + FN}$	$\underset{[0,1]}{Correctness} = \frac{TP}{TP + FP}$
Settlement Areas	84%	78%
Forest Areas	70%	96%
Road Network (Main Road)	69%	94%

Table 1: Quality Measures

It should be noted that these global quality measures are not able to detect details like the refinement of the road network by means of the extracted vegetation objects. Although the correctly extracted length of rows of trees was about 7[km], the improvement of the road network was only 1%. The main reason for this small increase is that rows of trees occur in different local context, e.g., with roads, creeks, or simply as a field border. Another reason is that in some cases if the road is only partially occluded by trees, the road had already been extracted in the first place.

7 SUMMARY AND OUTLOOK

Our work deals with the automatic interpretation of aerial imagery for GIS update. The focus of the research is on road and vegetation extraction and their mutual support. The paper is subdivided into three main parts. The first part describes the strategy which is based on global and local context knowledge and the reduction of the complexity of the image data. First it is reduced by means of scale space transformations for the extraction of global context knowledge. Then the complexity is semantically reduced by means of the global context. In the second part of the paper the model for the global context classes, their extraction from the coarse scale imagery, and the application of the global context knowledge is described. We also explore the relations of the abstraction of the image content and a GIS data model. The extraction of roads and trees and their mutual support in fine scale based on a common model is the topic of the third section. It also includes an external evaluation of the obtained results.

The explicit representation of the global context knowledge can be looked upon as a frame for the interpretation of a whole scene. The necessary global context for the extraction of objects not treated in our work is also given, e.g. the context *settlement* for building extraction. By means of the example Row of Trees, which originates from the used feature class catalogue we show a relatively direct conversion from a GIS data model into a model for object extraction and the application of this model. The strong link between the two models can be looked upon as a way towards a more explicit description of how maps are made (*).

Further work has to be done in the refinement of the tree model in order to improve the completeness. Another point is the integration of a topographic object which looks quite similar to a road with neighbouring trees – watercourses. In the actual implementation they were not considered in detail, but they should be introduced into the model because they are valid topographic objects.

ACKNOWLEDGEMENT

Parts of this work were developed within the IST Programme CROSSES financed by the European Commission under the project number IST-1999-10510.

REFERENCES

Baumgartner A., Eckstein W., Mayer, H., Heipke, C., Ebner, H., 1997,Context Supported Road Extraction, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser Verlag Basel, pp. 299-308.

Ebner H., Eckstein W., Heipke C., Mayer H., 1999, Editors, Proccedings of ISPRS Conference "Automatic Extraction of GIS Objects from Digital Imagrey", International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 3-2W5, Munich, September 8-10.

(*) “The problem of making maps automatically must first be solved on the conceptual level, however. As strange as it may be, but we do not know exactly how to make maps. That is to say, we cannot explicitly describe the map making process.” [SCHENK, 1993]

Förstner W., Liedtke C.-E., Bückner J., 1999, Editors, Workshop on "Semantic Modeling for the Acquisition of Topographic Information from Images and Maps" (SMATI'99), 7. Sept. 1999, Munich, Germany.

Fritsch, 1997, Experiences with the airborne three-line camera system DPA, Photogrammetric Week'97, Wichmann Verlag, Eds. Fritsch/Hobbie, pp. 63-74.

Gougeoun 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, Vol. 21, No. 3, pp. 274-284.

Gülch E., Mayer H., 1999, PE&RS Special Issue - Image Understanding – Introduction, Photogrammetric Engineering & Remote Sensing, Vol. 65, 7, p. 767.

Heipke C., Straub B., 1999a, Towards the automatic GIS update of vegetation areas from satellite imagery using digital landscape model as prior information, IAPRS, Vol. 32, Part 3-2W5, "Automatic Extraction of GIS Objects from Digital Imagery", München, Sept. 8-10, pp. 167–174.

Heipke, C., Straub, B., 1999b, Relations between Multiscale Imagery and GIS Aggregation Level for the Automatic Extraction of Vegetation Areas, Proceedings, ISPRS Joint Workshop, "Sensors and Mapping from Space 1999", September 27-30, Institute for Photogrammetry and Engineering Surveys, University of Hanover.

Kunz D., Vögtle T., 1999, Improved Land Use Classification by Means of a Digital Topographic Database and Integrated Knowledge Processing, Proceedings of the IEEE 1999 International Geoscience and Remote Sensing Symposium "Remote Sensing of the System Earth - A Challenge for the 21st Century", 28 June - 2 Juli 1999, Hamburg.

Lindeberg, T., 1996, Scale-Space: A framework for handling image structures at multiple scales, Proc. CERN School of Computing, Egmond aan Zee, The Netherlands, 8-21 September 1996, http://www.nada.kth.se/~tony/cern-review/cern-html/cern-html.html, (20 March 2000).

Lindeberg, T., 1994, Scale-Space Theory in Computer Vision, Kluwer Academic Publishers, Boston USA.

Mayer H., 1998, Automatische Objektextraktion aus digitalen Luftbildern, Deutsche Geodätische Kommission, Reihe C, Nr. 494.

Mayer H., 1999, Automatic Object Extraction from Aerial Imagery - A Survey Focusing on Buildings, Computer Vision and Image Understanding (74) 2, pp. 138-149.

Pinz A., 1989, Final Results of the Vision Expert System VES: Finding Trees in Aerial Photographs, Pinz (ed.), Wissensbasierte Mustererkennung (Knowledge Based Pattern Recognition), OCG-Schriftenreihe Oldenbourg. pp. 90-111.

Rapp F., 1995, Modell und Realität, Zeitschrift für Photogrammetrie und Fernerkundung, 6/1995, pp. 220-223.

Richards J.A., Jia X., 1999, Remote sensing digital image analysis, Springer, Berlin.

Schenk T., 1993, Image Understanding and Digital Photogrammetry, Photogrammetric Week'93, Fritsch/Hobbie (ed.), Karlsruhe, Wichmann, pp. 197-207.

Schiewe J., 1999, An Advanced Technique for Pixel-Based Multi-Sensor Data Integration, Proceedings, ISPRS Joint Workshop, "Sensors and Mapping from Space 1999", September 27-30, Institute for Photogrammetry and Engineering Surveys, University of Hanover.

Space Imaging, 2000, IKONOS, http://www.spaceimaging.com/aboutus/satellites/IKONOS/ikonos.html, (4 March 2000).

Suetens P., Fua P, Hanson A.J. , 1992, Computational Strategies for Object Recognition, ACM Computing Surveys, Vol. 24, No. 1, pp. 5-60.

Steger C., 1998: An unbiased detector of curvilinear structures, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20 No. 2, pp. 113–125.

Wald L., Ranchin T., Mangolini M., 1997, Fusion of Satellite Images of Different Spatial Resolutions: Assessing the Quality of Resulting Images, Photogrammetric Engineering & Remote Sensing, Vol. 63, No. 6, pp. 691-699.

Wiedemann C., Heipke C., Mayer H., Jamet O., 1998, Empirical Evaluation of Automatically Extracted Road Axes, Kevin J. Bowyer and P. Jonathon Phillips, (ed.), Empirical Evaluation Methods in Computer Vision, pp. 172–187, Los Alamitos, California, IEEE Computer Society Press.

Wiedemann C., 1999, Completion of automatically extracted road networks based on the function of roads, International Archives of Photogrammetry and Remote Sensing Vol. 32(3-2W5), pp. 209-215.