

A FLOWCHART INTERFACE FOR HYBRID ANALYSES IN AN INTEGRATED GIS

Stefan Jung, Stefan A. Voser and Manfred Ehlers
ISPA, University Vechta
P.O. Box 1553, D-49377 Vechta, Germany
(jung, voser, ehlers)@ispa.uni-vechta.de

Commission II, Working Group 2

KEY WORDS: Hybrid Analysis, Spatial Analysis, Integrated GIS.

ABSTRACT

For non-GIS experts the complex analytic capabilities of most Geographic Information Systems are hard to handle, especially if different data sources such as cadastral data or remotely sensed data are involved within the same task. The Virtual-GIS (VGIS) intends to offer relief regarding different shortcomings of existing systems.

A workflow like interface enables the user to design complex spatial analysis processes onscreen and combines task design and actual processing in an ideal way. The Graphical User Interface (GUI) can be addressed as a visual programming environment.

Workflow applications can be built upon a set of 20 universal GIS operations that work independent of data structure and are therefore able to deal with different data types (raster, vector).

The design necessary to realize truly hybrid high level operators is outlined and a step towards specifying interfaces of these operators is made.

To develop hybrid control mechanisms, the semantics of the data have to be considered because raster data may describe raw satellite imagery, interpreted raster data, terrain models etc. Depending on the semantic of input data hybrid operators have to be able to perform different operations.

At the highest level of abstraction, the users perception, a reasonable control can only be realized through the use of metadata. A concept for metadata management is offered distinguishing between metadata for data and metadata for operators.

Finally the hybrid analysis capabilities of the universal operators are analyzed and requirements for their technical realization are depicted.

1 INTRODUCTION

GIS or environmental analysis tasks are specific to Information Communities. They are often well defined and raise unique spatial questions. All spatial data analysis systems, including image processing and GIS offer a large amount of algorithms and analytical operations to solve spatial problems. For non-experts, this variety as well as the complexity of operations to perform which is mostly technical in nature impedes fast digital solutions.

Necessary system decisions, different data formats, data structures, and data models as well as lack of compatibility are additional drawbacks to GIS employment for different kinds of spatially related tasks. On the other hand, a broad technical toolbox is wanted that is capable of supplying full functionality for a wide range of freely defined problems.

Consequently there are strong demands to design operators closer to the user, but still universal, which means independent of the Information Community, who wants to use them.

Embedded in a flowchart based Graphical User Interface (GUI) a set of 20 universal GIS operations has been created to fill the gap between the great variety in analytical GIS functions and their usability. This solution addresses several aspects necessary to guarantee the ease-of-use of an analytical GIS-Tool (Virtual GIS or VGIS) (Albrecht 1996, 1997). A major step towards the construction of a truly hybrid (i.e. integrated) analysis tool is the hierarchically structured operator design containing approaches towards a hybrid analysis.

During the process of developing an analytical solution for application specific spatial questions within the VGIS tool, a sequence of different operators is created that can be regarded and treated as a specific form of a graph. The visual programming potential of VGIS makes these process-graphs executable similar to a graphical case tool. To assure consistency within such a graph, VGIS needs a *data control mechanism* through the use of metadata and a *process control mechanism*.

The design of a generic universal high level GIS operator is described in section 2.

Specifications are lined out which will be needed to realize interaction between different operators (section 3) before an approach towards metadata management is offered in section 4 and underlying mechanisms to realize hybrid processing will be explained (section 5).

The conclusion will focus on further research topics in the future.

2 THE DESIGN OF GIS-ANALYSIS-OPERATORS

The design of Analysis-Operators in GIS covers structure and behavior. Behavior has to be defined as the abstract function performed by the operator, independent of data structure.

Structure describes the way an operator is designed to perform data structure independent processing. An operator is able to analyze input data, to select an adequate operation depending upon the kind of input data, and to return output data. Inside the operator, data driven selection processes ensure that adequate algorithms are chosen to analyze the specific input data types.

To understand what happens inside an operator, the view of a multi-shell object just like an onion comes into one's mind, symbolizing to a certain degree a hierarchical structure. In case of universal operators, we see three different levels:

The *management level* is the outmost shell the user interacts with. At this level, a certain problem of an information community is tackled through the knowledge and experience of the user. The medium for the problem solution is offered in form of the above GUI with visual programming capabilities and universal operators. Users can use stored graphs, create new graphs, change and assign parameters to the operators and create their personalized solutions. Communication takes place through the use of metadata in data catalogues and the operators. Metadata are passed through the chain of activated operators.

Semantic of an operator is realized at the *control level* (mid level) where input data, metadata and the supplied parameters are analyzed and translated into the adequate processing chain (polymorphism).

Processing then is performed on a very low level (*processing level*), depending upon the distributed computing platform, operating system, and the software environment.

Management (High Level): At high level, the semantics of the operation is defined. The data to be analyzed are selected from the data catalogue with their associated metadata. The operator is specified upon its parametrization which is also metadata driven. This level manages the whole operator. It is the constituent part of the analysis task. The user defines the operation upon his cognitive experience.

The data description at this level is independent of any spatial data structure (e.g. raster or vector) or data schema.

Controlling (Mid Level): At mid level, the controlling of the operation is carried out. The input data, selected at high level, are linked to their corresponding data representations in a geometric and structural level. The operator analyzes the metadata, given at the management level, and chooses the algorithm that matches the data types and other requirements. Here, the requirements for hybrid analysis functionality arise because of the different data representations.

At this level, the cognitive semantics of the data and operation gets lost, the information is translated into syntactically structured information for the processing in which their cognitive context is unimportant.

Processing (Low Level): The low level is the processing level. The algorithm processes the data numerically, all operation's characteristics are processed and sent back to mid level.

A universal analytical GIS-operator is characterized by its functionality. It is important to look upon this functionality in the context of input and output data (see *Figure 1*).

The universal operator is able to analyze the input data to choose the appropriate algorithm for processing. As a result, new output data are generated. A High-Level-Operator therefore has to be metadata driven. For selecting data and functionality, only metadata will be used. The following survey introduces the concept of operators and their input and output data.

structures of the input data. Each algorithm has its own profile. The process control executes an analysis of the data as an inquiry for the right algorithm. The operator is designed in a polymorph way which includes hybrid operations.

Output data are mapped into a data catalogue which may already exist. Otherwise a data catalogue including metadata and lineage information has to be generated. Its representation is linked to the underlying data schema. The output data catalogue may be derived from the operator, or the operator has to meet the conditions of the data model chosen by the user.

The functionality of a GIS-Analysis-Operator can be divided in the three categories *geometric converter*, *semantic translator*, and *metadata processor*.

All geographic data have a *geometric* component. One of the main goals of spatial analysis is to solve geometric (metric and topological) questions. Generally a GIS analysis produces data with new geometric information. Consequently a GIS-Analysis-Operator converts the geometric input into the new geometry of the output data.

Semantic of data is assigned by the Information Community on the basis of a conceptual model with attributes and descriptions. It is located above the underlying geometry. The geometric process of the operator is linked to the thematic component of the data, a new theme is generated during the combination process with the geometric and semantic overlay. The result is a new semantic.

The Operator is controlled and prepared for processing, based on the *metadata* of the input data. Metadata is translated to process control parameters. The process produces new metadata. The generation of the output metadata is controlled by the operation's metadata.

The controlling of the operator includes the preconditions that are derived from the input data and the chosen operation with its implemented algorithms. For the output, the postconditions include the requirements of the output data. The operation has to be protocolled including the lineage and other metadata (Jung and Albrecht 1997).

3 SPECIFICATION OF HIGH LEVEL OPERATORS

Universal operators interact with different instances. Existing specifications, especially those that are about to become standards, have to be regarded and are at the same time a valuable foundation for the definition of universal high level operations. As a result for well structured and designed operators formal specifications will emerge.

Specifications of the OGC will have to be considered for the implementation of features, coverages, data catalogues and metadata (OGC 1996; OGC 1996). Specifications are also needed for the user interface of operators.

Based upon the operator design described above, the specification of a generic high level spatial analysis operator (Meyer 1997) is necessary to create instances of an operator class "high level-operators". An instance of that class could be any high level operator, e.g. buffer, overlay, shortest path ... At the current state of development the specification introduced in this section does not claim to be complete or formal in mathematical terms. It rather serves as a more detailed description of components in order to approach a formal specification. It is the attempt to outline the general conditions necessary to create high level operators. Therefore it has to be regarded as a step towards the realization of operator design.

The variety of instances of high level operators, design and specification comprises different aspects. The data structure

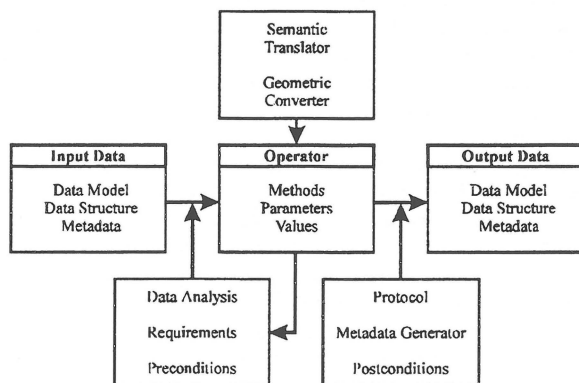


Figure 1: Operator Design

Input data are described through a data catalogue including its metadata and lineage information. The link to their digital representation in the related and implemented data schema has to be known.

The *operator* is defined by a generic method. This is specified by its characteristic parameters. The operation is described by the operator's metadata. The operator generates specific metadata as a protocol for the lineage. The underlying control has to select the correct algorithm that fits the types and

independence of the management level has to be implemented through a data type driven polymorphism at the control level. Due to the polymorphism a high level operator is able to execute different algorithms or sequences of algorithms at the processing level.

In addition, cognitive and semantic aspects can not be neglected in operator design. Yet the structure of data plays a central role in the building process of high level operators concerning different parts of the specification.

3.1 User Interface

The user interface specifies what kind of interaction can take place between the user and the operator. User interaction includes any form of communication between the user and the operator, such as:

<i>operator control</i>	Selecting data through a data catalogue Input or determination of parameters for the operator.
<i>user support</i>	<i>documentation</i> Access to documentation and description (metadata) of operations and data. Transparency of the operator by describing architecture and algorithms.
<i>process control</i>	<i>messages</i> User comfort with messages and warnings Error-control.

3.2 Plausibility Control

Plausibility control of an operator should be able to determine whether an operation proves to be meaningful or not. Representation of this advanced form of semantics has to be one level above the creation of the correct data type management. It comprises several aspects.

<i>Semantics Control</i>	Analysis and comparison of semantics in input and output data should enable the operator to distinguish between allowed and forbidden combinations. Predefined semantic results for given input/output combinations could also be attached to the operator (<i>semantic templates</i>).
<i>Property Control</i>	A relation between geometric and semantic properties of data should be established. For example, geometric accuracy control should decide which combination of different resolutions is allowed. Tolerated ranges of accuracy for different semantics of spatial data can be supplied.
<i>Lineage Control</i>	Lineage information attached to output data could include plausibility control documentation and results.

3.3 Data Types

Assumed that spatial data are organised in *data types*, it has to be exactly specified which data types can be processed by an operator and which data types will be produced by the operator after processing. Data types distinguish between different representations of geometry within spatial data.

<i>Catalogues</i>	Specification of allowed data types and its definitions. data model (topological vector model, task
-------------------	--------------------------------------------------------------------------------------------------------

specific models, ...).

structures (raster, vector, tin ...).

Combinations List of input-output-couples (combinations of input and output data types related to an operator).

Restrictions Restriction for operator use with regard to the allowed *data couples*.

3.4 Polymorphism

Polymorphism stands for different algorithms related to different data types of the same high level operator. Program structures are created at the control level by decisions made according to data types, user parametrisation and semantic specifications. Certain conditions have to be met:

<i>Uniqueness</i>	Non-ambiguous decision rules.
<i>Completeness</i>	A function must exist for every combination of allowed input data.
<i>Correctness</i>	A method to test consistency and correctness of decision rules.
<i>Extensibility</i>	Operations at control level must be extensible to new data types and new algorithms.

3.5 Algorithms

At processing level standard software development guidelines have to be regarded. It is planned to choose an object-oriented approach with certain advantages related to the described operator design.

<i>Transparency</i>	Transparency of algorithms by documentation.
<i>Uniqueness</i>	Avoiding identical processing and geometric results for different high level operators using the same algorithm.
<i>Redundancy Free</i>	Non-redundant implementation of algorithms (modules, function libraries, etc.).

3.6 Application Specification

Optional is an extension of high level operators towards the restricted and/or expanded use in certain information communities. That would require:

<i>Limitations or Reductions</i>	Restriction in functionality according to a reduction of defined operations in addition to the specifications mentioned so far.
<i>Specialization</i>	Specialization of certain operators through the input of domain specific data and functionality (e.g. integration of domain specific rules and algorithms).

4 METADATA

Metadata is used to manage all operators at the highest level of abstraction. At the interface or management level, only metadata control user interaction. Explicit and implicit metadata are analysed to control the operator at mid level. The metadata have different meanings at the three levels as described in chapter 2.

In the following, the processing of metadata for high level-GIS-operators is divided into two parts: metadata of spatial data (Ganter 1993) and metadata of the operator.

4.1 Metadata of Data

The metadata describe the data at a high level of abstraction in a data catalogue with its related information. The information is the link to the database and to its data schema in which all other information is stored implicitly.

- Management** (High Level) The metadata of the data describe the content and the conceptual organisation of the data.
- Controlling** (Mid Level) At mid level, the semantics of the metadata of high level has no meaning. The information of the data types, in which the data are represented, mainly important for the control of the respective polymorphism. The main aim at mid level is the control of the structural information of input and output data. This includes the control of the polymorphism of the operator.
- The operator transforms the semantics of the metadata to parameters used for processing, whereas the semantics of the data is of no relevance for the parameters of a process.
- Processing** (Low Level) At low level, metadata focus on the values and categories of the parameters. The characteristic parameters of the operators and its values are instantiated and assigned to the data to be processed.

4.2 Metadata of Operators

The operator metadata control their spatial analysis process. They map and process the metadata of input and output and protocol the operation and the lineage.

- Management** (High Level) The metadata of the operators describe the task, the functional behavior and the required information. The operator's metadata include the semantics, which is related to the data.
- Controlling** (Mid Level) The operator's metadata control the operation. At this level the mapping of the data types to the algorithm of the polymorph implementation takes place.
- Processing** (Low Level) This level generates the input data for the lineage of the operation which is related to the derived data.

5 HYBRID ANALYSIS

Hybrid analysis is needed for the integration of GIS and remote sensing data (Ehlers 1993; Woodsford 1994; Hinton 1996; Wilkinson 1996), for terrain analysis (DTM's in raster format with overlay of vector data) etc.

Within VGIS hybrid analysis is enabled through the polymorphism of the universal operators. Depending upon the structure of input data different algorithms can be invoked to perform requests as well as the creation of new geometries and objects. Emphasis has to be placed upon the operations that combine input data-types with different data structure (e.g. raster with vector objects, raster with TIN objects etc.).

Possible results of a hybrid analysis are new raster data, vector data or both of them.

Some examples of hybrid analysis functionality

- Fencing off raster processes by the overlay of vector data.
- Verification of geometry and attributes.

- Transferring extracted geometry from raster to vector.
- Transferring attributes from image interpretation to vector data.
- Transferring attributes form vector databases to raster data
- Using raster information for determination of uncertainty of vector data.
- Equalizing Data.

The complexity of the hybrid functionality may be shown by the following survey:

- Projecting Information** geometric overlay.
semantic overlay.
- Combined Analysis** derivation by aggregation.
derivation by interaction.
accumulation by information transfer.

Knowledge-based Conversion knowledge-based conversion
extraction of objects.

The full range of spatial analysis may be reached with operations that interact with raster and vector data without the need of conversion. The requirements of hybrid analysis are divided in geometric interaction, semantic analysis and metadata processing (Egenhofer 1993).

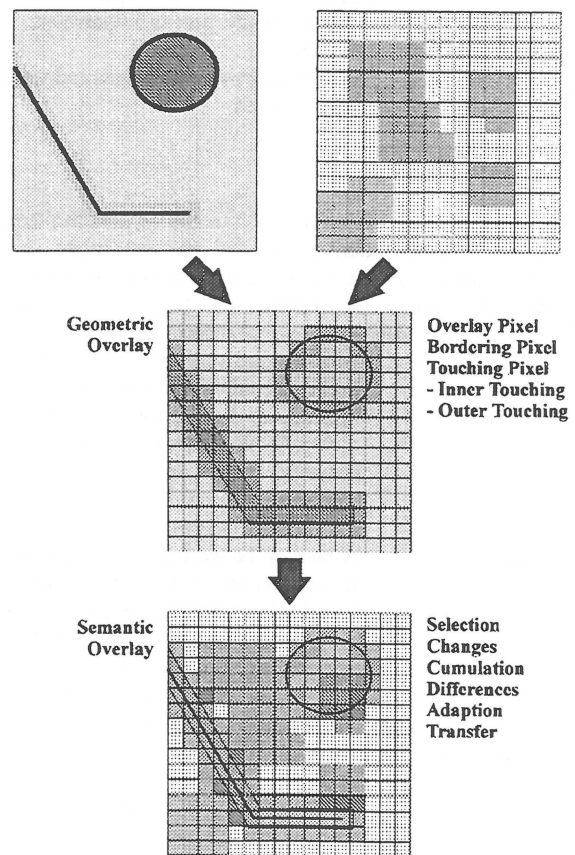


Figure 2: The hybrid principle

Geometric Analysis *positional location*: generating the relation of coordinates of different data and identifying identical positions of geometric primitives.
Linking geometric features: identifying features covering the same location.
Extracting and transferring geometry: generating new geometry by extraction, transfer and interpolation.

Semantic Analysis *geometric interpretation*: giving new semantics to the data.
Thematic analysis: linking, projecting and deriving integrates new information.
Combination of geometric and thematic interpretation: combined analysis of geometric and thematic information.

Metadata Processing analysis of metadata.
Description of the geometric, semantic and combined analysis.
Description of the semantic analysis.
Statistics about geometric overlay.

6 CONCLUSION

The concept of universal GIS analysis operators independent of Information Communities is outlined and specified. Universality in this context includes a domain independence as well as the option to work with different data structures. This can only be achieved through metadata driven user interaction. To realize the data structure independence of universal operators, hybrid analyses have to be integrated. Through a currently increasing demand for the integration of remote sensing and GIS, hybrid analysis techniques have gained significance. Our future work will focus on a detailed specification of hybrid analysis operations and their prototypical implementation.

7 REFERENCES

- Albrecht, J. (1996). "Universal Analytical GIS Operations: a task-oriented systematization of data structure-independent GIS functionality leading towards a geographic modeling language." Dissertation ISPA Mitteilungen University of Vechta: Vechta, Germany. , 23.
- Albrecht, J., S. Jung and S. Mann (1997). VGIS-a GIS Shell for the Conceptual Design of Environmental Models. Innovations in GIS, Taylor & Francis. 4, pp. 154-165.
- Egenhofer, M. J. a. S., J. (1993). "Topological Relations between regions in R2 and Z2,." Advances in Spatial Databases -Third International Symposium, SSD '93 Lecture Notes in Computer Science 692, pp. 36-52.
- Ehlers, M. (1993). "Integration of GIS, remote sensing, photogrammetry and cartography: the geoinformatics approach." Geo-Informations-Systeme 6(5), pp. 18-23.
- Ganter, J. (1993). Metadata Management in an Environmental GIS for Multidisciplinary Users. Proceedings GIS/LIS '93,.
- Hinton, J. (1996). "GIS and Remote Sensing Integration for Environmental Applications." International Journal of Geographical Information Systems 10(7) pp. 877-890.
- Jung, S. and J. Albrecht (1997). Multi-Level Comparative Analysis of Spatial Operators in GIS and Remote Sensing as a Foundation for an Integrated GIS. Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, SMATI '97. W. u. L. P. Förstner, Birkhäuser Verlag, pp. 72-88.
- Meyer, B. (1997). *Object-Oriented Software Construction*. New Jersey 07458, Prentice Hall PTR.
- OGC (1996). "The OpenGIS Abstract Specification: an Object Model for Interoperable Geoprocessing,." Open GIS Consortium:Revision 1. OpenGIS Project Document Number 96-015R1.
- OGC (1996). The OpenGIS Guide: Introduction to Interoperable Geoprocessing, OpenGIS TC Document Number 96-001.
- Wilkinson, G. (1996). "A Review of current Issues in the integration of GIS and remote sensing data." International Journal of Geographical Informations Systems, 10(1), pp. 85 - 101.
- Woodsford, P. (1994). "Integration of Remote Sensing and GIS,." 30(2), pp. 383 - 390.