HYBRID SPATIAL ANALYSIS OPERATIONS AS A FOUNDATION FOR 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

Keywords: Hybrid Analysis, Spatial Analysis, Integrated GIS

ABSTRACT

In many applications, especially when remote sensing data are needed, both vector and raster data have to be processed in a GIS. Usually this takes place through different conversion methods after the user has decided if he wants to work in the raster or vector world, depending upon the task he wishes to fulfill.

With the use of an existing definition of spatial relations between objects, based on a mathematical framework, the foundation for a taxonomy of hybrid operations is laid. This taxonomy shall guarantee that different queries, comparison and overlay operations will be integrated in the 20 universal operations of the Virtual GIS (VGIS). In VGIS spatial data of different structure can be processed in combination with a flowchart like GUI to perform spatial analysis.

The core of the hybrid operations are the rules that formally define how different data structures are interpreted in combined queries (hybrid analysis). An effective hybrid functionality can only be achieved through the use of a minimal set of metadata.

1   INTRODUCTION

GIS or environmental analysis tasks are specific to Information Communities. Often actuality plays an essential role in (environmental) applications and in many cases data of large areas are needed. Remotely sensed data are likely to be the most recent information sources available and satellite data are even regularly produced.

The data model as well as the data structure of digital aerial photographs and satellite images or scanned maps and images significantly differ from the object oriented view mostly realized with vector data structure. Applications are asked for that can process those different kinds of data in spatial analysis at a similar level of abstraction.

Following these demands, an Integrated GIS (IGIS) needs more sophisticated functionality then mere structural data conversion algorithms. A set of basic and universal hybrid operators shall be created which are build mainly upon well known, existing query and analysis methods.

The Virtual GIS (VGIS) supplies a Graphical User Interface as frontend to embed universal hybrid operators in a workflow-like visual programming environment (section 3).

Before offering ideas defining complete and therefore complex operators, a distinction has to be made between different data structures and data types of geodata (section 2). A coarse taxonomy of possible hybrid operations (section 3) will then be derived from the thoughts concerning data structures and types.

Further steps towards a technical realization of hybrid processing are suggested in the geometric overlay of raster and vector data (section 5) and their mathematical framework (section 6). Finally some thoughts about metadata are presented (section 7) before in a conclusion the focus of future work is shown.

2   DATA STRUCTURES AND DATA TYPES

Data of real world phenomena will always stand for an abstraction of reality according to a model from a particular, application specific point of view. For digital processing each real world model is represented in a specific *data structure*. This structure either is best suited to the model chosen or it goes back to the data acquisition technique used. Landsat/TM images for instance are always delivered in the same structure. Different GIS also have certain proprietary data structures for different applications (Peuquet 1991).

An essential part of the data structure especially in the context of spatial applications is the *geometric representation*, e.g. vector (point, line, area) and raster (pixel, grid). Hybrid spatial analysis is meant to control interaction between vector and raster primitives (Voser, Jung et al. 1998). Managing and comparing these different geometric representations is a core functionality of each hybrid operation.

The term *data type* goes beyond pure structural characteristics and includes cognitive and semantic aspects of the data model. The differentiation in data types can be understood as a classification of abstraction models (Gahegan and Flack 1996) supplying semantics in addition to geometric structure.

The problem of hybrid analysis can be broken down to a set of rules how to treat different data types. Abstract functions with relation to space need to be implemented differently depending upon the geometrical representation of the data (raster, vector, ...).

To create a taxonomy of hybrid operations, abstract functions have to be defined regarding these differences in data types (Jung and Albrecht 1997).

D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

Jung et al.   277

3 VGIS, THE WORKFLOW TOOL TO PERFORM COMPLEX GIS ANALYSIS

The Virtual GIS (VGIS), currently at a prototypical state, functions as user interface and integrated environment for the implementation of hybrid operations.

VGIS provides a visual programming environment that functions like a graphical macro language. With drag and drop techniques operators can be pulled on the VGIS canvas where they are to be arranged to design individual process flows. The operators can then be linked to create executable workflows, that sequentially perform spatial analysis steps. In Figure 1 the example of a spatial decision process concerning the optimal location of a chemical factory has been realized with a VGIS-flowchart. When the workflow is executed all connected operators perform their action in sequence.

These workflows can be stored and managed as files. Workflows can also be designed and stored as compound operators which can then be integrated in other workflows.

The principal idea behind this kind of GUI is to build reusable and easily managable process libraries for different applications (Albrecht, Brösamle et al. 1996.).
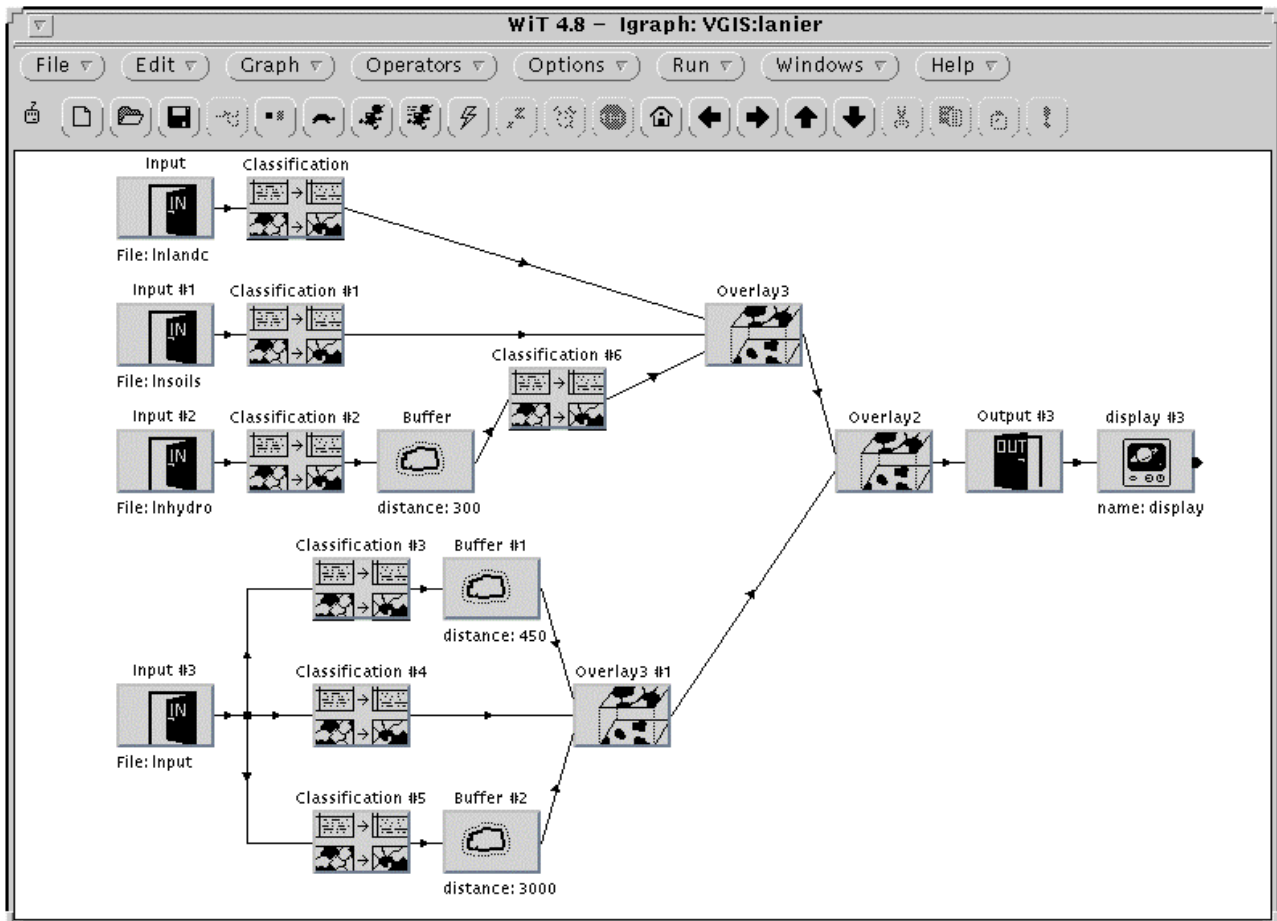

Figure 1: Flowchart example in VGIS

To be able to construct process flows, 20 universal GIS operators grouped in six classes (Figure 2), are offered. These operators are designed to meet certain requirements to make them universal (Albrecht 1996):

- they should be data structure independent, that means applicable to raster as well as to vector data sets.
- they should not depend upon certain domain specific applications, e.g. there should not be any explicit planning operators.
- their functionality is limited to purely analytical functions.
- the operators shall cover the complete range of spatial analysis.
- the design of the operators, consisting of naming and description, should be self explanatory for the inexperienced user.

A representative subset of these 20 universal operators have been implemented for a raster and a vector system to function in a quite similar manner. Until now the combined spatial analysis of raster and vector data yet has not been realized in detail. There is an urgent need for a complete set of rules to combine those different geometrical representations.
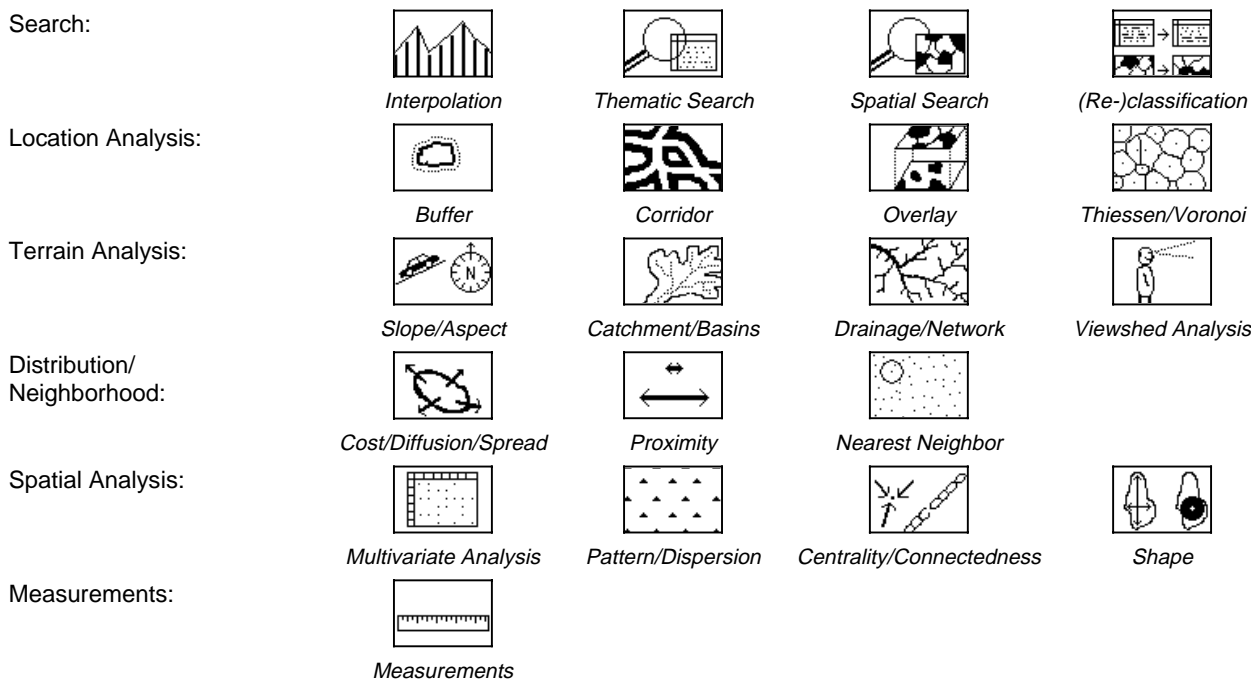
D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

278   IAPRS, Vol. 32, Part 4 "GIS-Between Visions and Applications", Stuttgart, 1998

Search:

Interpolation *Thematic Search* *Spatial Search* *(Re-)classification*

Location Analysis:

Buffer *Corridor* *Overlay* *Thiessen/Voronoi*

Terrain Analysis:

Slope/Aspect *Catchment/Basins* *Drainage/Network* *Viewshed Analysis*

Distribution/
Neighborhood:

Cost/Diffusion/Spread *Proximity* *Nearest Neighbor*

Spatial Analysis:

Multivariate Analysis *Pattern/Dispersion* *Centrality/Connectedness* *Shape*

Measurements:

Measurements

Figure 2. The 20 universal analytical GIS operations

4 TAXONOMY OF HYBRID OPERATIONS

In establishing a complete taxonomy of hybrid operations (in form of abstract functions) a reasonable first step is to extract the data type sensitive interaction between different data sets. The difference in data structure between raster and vector shall be sufficient to demonstrate the problem.

An approach towards hybrid GIS analysis operations can be broken down in two classes of operations.

"Simple" operations concern functions that are applied to one data set consisting of only one data type, e.g. buffering, measurement, reclassification, selection processes ... In this case emphasis has to be put upon a formal description of the semantics of an operation to achieve a performance as similar as possible when applied to different data types (different geometric representations). For example, a buffering applied to a selection of a raster data set should yield nearly the same result as the comparable operation performed on vector data.

A formalized object oriented approach of an abstract data type as generic operator, with different implementations for each data type seems to be an appropriate way to realize this concept.

Hybridity becomes more complex when different data types are to be combined in spatial and/or thematic analysis. Especially raster data, but to a certain extent vector data as well, need additional information in form of metadata to be interpreted correctly (see section 7).

Analysis of data with different geometrical representations needs processing rules as soon as an overlay of geometry is involved. Geometric Overlay occurs within different functions extracting and generating geometric and/or thematic data. They can be aggregated to:

Combined Queries:
 one data set (data type a) shall return information according to a selection set of geometries in another data set (data type b).

In this case rules are needed that define in what case which kind of information is returned, but no change in geometry occurs (Wilkinson 1996).

Comparison/Verification:
 Comparison of the geometry of different geometric representations (Raster - Vector) with rules describing tolerances, form parameters etc. The data of one data set influence the data of the other one.
 The goal is a set of criteria which enables for instance the assignment of (exact) feature geometries from existing vector data to extraction results from image interpretation processes.

Overlay:
 new features with new geometry are created through the combination of thematic attributes and geometry of different data sets.
 Principally the Comparison/Verification rules can be used to decide when and where to create new features.

Complementary the semantics of data types have to be regarded in relation to the meaning of geometry. The kind of scale of raster cell values for instance informs the system (or the operator) if interpolation between values is allowed or if the value has a mere descriptive function.

5 GEOMETRIC OVERLAY OF VECTOR AND RASTER DATA

Traditionally any form of combined analysis between different data types, especially between vector and raster representations, involves conversion between them (Congalton 1997). The hybrid approach tries to avoid explicit conversion wherever it is possible. Yet strictly formalized rules need to be defined to combine the two different geometrical representations.

D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

Jung et al.   279

Again different cases can be separated, this time according to the semantics of the pixel values. Mainly the two different views of spatial data, the *field view* and the *object view* can be distinguished (Couclelis 1992).
Field view data as a continuous raster data type consists of point values located at a defined location within the pixel. The space between the explicit point locations of the data set is continuously and completely covered by the pixel values and a defined interpolation algorithm delivered. This data model is often used to describe phenomena continuously varying over space and/or time as e.g. digital terrain models, precipitation, climatic parameters ... Field view data, organized as described, allow any kind of interaction with vector data since at each possible point in the **R²** space is defined through the pixel values and interpolation rule.
A similar process has to be applied to *object view* data. In addition to the raster values itself, rules must define how to interpret them, meaning how to treat them in possible analysis processes. Well known approaches (Tomlin 1990) show how raster information can be adequately interpreted even without additional data, just taking into account the distribution (pattern) of raster cell values over space.

Raster cells may represent objects of different dimensions in the same data structure (Point objects, line objects or area objects). In most cases, the data set is not able to say anything about the dimension of included objects without additional information (metadata).
To show how geometric overlays between raster and vector could be managed, the restriction to 2 dimensional objects represented in raster structure will be sufficient. If each raster cell represents a part of an 2 dimensional object it could be regarded as a vector object with the extent of the raster cell as (vector) geometry and the value of the raster cell as attribute. Possible spatial relations between vector objects are well described and formalized in Egenhofers mathematical framework (Egenhofer and Herring 1990) and could now be applied to the raster cells in combination with all possible vector objects.
For example a spatial query, that would have to be performed over all raster cells - effective indexing not regarded at the moment - could be: Does a raster cell belong to a vector object ? (see Table 1). At this example no additional information is given, but simple unique query rules for the interaction of raster and vector structures.

Vector object	raster cell (= vector area)	possible relation	Query: Does a raster cell belongs to an object ?
point	area	within	yes
point	area	touches	yes
point	area	not within	no
line	area	within	yes
line	area	crosses	yes
line	area	touches	yes
line	area	not within	no
area	area	within (subtypes possible)	yes
area	area	not within	no
area	area	crosses	no
area	area	touches	no
area	area	contains	yes

Table 1:   Query results for an overlay query between raster and vector data sets

The result in this case could be a relational table between a set of vector objects and a set of raster cells in a n:m relation. Refinements can be thought of concerning the relations between objects as an extension of Egenhofer's nine intersection model. Different algorithms can be used to determine membership or non membership of a raster cell according to an overlay objects geometry (Congalton 1997).
To set up more sophisticated rules, the semantic of data has to be taken into account and metadata have to be added.

6   MATHEMATICAL REPRESENTATION

To guarantee a consistent query language regarding spatial relations between objects, Egenhofer offers in a first essential step mathematical representations of the **R²** space for vector data structures and a similar fundament regarding the **Z²** space for raster data (Egenhofer 1993). But this unification does not yet supply full compatibility between the different data structures.
The principal difference between raster and vector data structures is their spatial representation. While vector data are embedded in a continuous infinite space, the range of values of raster data is restricted to the discrete space of real numbers. Yet the meaning of the raster cell values is not limited to the specific location it is attached to, e.g. the

integer representing the pixels relative position. Additional information about the semantics of the data (stored in metadata) can tell how raster cell values have to be interpolated to fill the gap between them. Formally these additional information has to be precise enough (f:) to project the integer raster values (**Z²**) into a value for each real world coordinate pair (**R²**) (see (1)). This corresponds to the idea of the abstract coverage specification as part of the OGIS standardization development (OGC 1996). As result a request at any coordinate in space will retrieve some sort of value through the function applied to the raster data set (nodata value is allowed).

$$f: \mathbf{Z^2} \rightarrow \mathbf{R^2} \qquad (1)$$

A function f can be used to model continuous phenomena as well as discrete objects, depending on the type (semantics) of data.

7   METADATA

Metadata in this context concerns only those kind of additional information that is needed to perform hybrid analysis. Metadata for geodata in general, including projections, coordinate systems etc., is a topic for the work of various standardization organization (ISO 1996; OGC 1996).

D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

280   IAPRS, Vol. 32, Part 4 "GIS-Between Visions and Applications", Stuttgart, 1998

As shown above metadata is not obligatory for applying hybrid operations, but it will improve results tremendously. Especially in cases where the dimension of the objects within a data set is not obvious additional information is essential.

Metadata itself can consist of different types of scale in describing data. A few basic data to support hybrid analysis are listed supplying field name, semantic description, and the type of data field in a possible metadata table (Table 2).

Field name	Description	Field type
value	Description of what the raster cell values mean	Character, memo, description
type of scale	What type of scale do the raster cell values belong to	Choice of scale types (character)
field or object view	Yes/No	Boolean
dimension of objects	if objects are described, which dimension may occur	Integer values
interpolation algorithm	if continuous phenomena are described, how should the values be interpolated	Formula

Table 2: possible metadata model

8   CONCLUSION

A basic concept has been presented to integrate spatial data of different data models and data structures within one tool to perform truly hybrid analysis. At the user interaction level a set of 20 universal operations shall cover full GIS analysis functionality. Different kinds of interaction between raster and vector data have been shown in a hybrid taxonomy.

One major idea is to minimize data conversion between different data structures, especially if queries are demanded.

Existing concepts of space and different geometrical representations can be combined efficiently to enable enhanced spatial analysis and improved performance.

A minimal set of Metadata can be sufficient to retrieve much better results from combined raster vector queries. The ideas shall offer a frame for further work to fill the still existing gap between the different structure of spatial data.

In a next step, a more elaborate connection between the mathematical foundation and the potential that lies in the use of metadata is necessary to lead to an applicable taxonomy of hybrid rules, tested and confirmed with examples.

9   REFERENCES

Albrecht, J. (1996). *Universal Analytical GIS Operations. A Task-Oriented Systemization of Data Structure-Independent GIS Functionality Leading Towards a Geographic Modeling Language*. Vechta, University of Vechta, ISPA.

Albrecht, J., H. Brösamle, et al. (1996.). *A Graphical Front-End for User-Oriented Analytical GIS Operations.* Proceedings XVIII ISPRS Congress, XXXI, B2, Commission II,Vienna, Austria,: 78-88.

Congalton, R. G. (1997). *Exploring and Evaluating the Consequences of Vector-to-Raster and Raster-to-Vector Conversion*. Photogrammetric Engineering & Remote Sensing **63**(4): 425-434.

Couclelis, H. (1992). *Beyond the Raster-Vector Debate in GIS.* Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. A. Frank, Campari, I. and U. Formentini (Eds.). Berlin, NewYork, Springer**:** 65-77.

Egenhofer, M. J. and J. Herring (1990). *A mathematical framework for the definition of topological relationships*. Proceedings of the Fourth International Symposium on Spatial Data Handling: 803-813.

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**: 36-52.

Gahegan, M. and J. Flack (1996). *A Model to Support the Integration of Image Understanding Techniques within a GIS*. Photogrammetric Engineering & Remote Sensing **LXII**(5): 483-490.

ISO (1996). *ISO 15046-15 Geographic Information - Metadata - Version 1.0.* http://www.statkart.no/isotc211/.

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**:** 72-88.

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*.

Peuquet, D. J. (1991). *A conceptual framework and comparison of spatial data models*. Introductory readings in geographic information systems. D. J. a. D. M. Peuquet .

Tomlin, C. D. (1990). *Geographic Information Systems and Cartographic Modeling*. Englewood Cliffs, Prentice Hall.

Voser, S., S. Jung, et al. (1998). Towards Hybrid Analysis - Specifications of High Level Operators. AGILE, Enschede, Holland.

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): 85 - 101.