ON REPRESENTING GEOMETRIES OF GEOGRAPHIC SPACE

Thomas Bittner and Andrew U. Frank
Department of Geoinformation, Technical University Vienna
{bittner,frank}@geoinfo.tuwien.ac.at

Keywords: Geometry, Location, Representation, Abstraction

ABSTRACT

This paper discusses the representation of phenomena of geographic space by means of geometry. It is argued that different kinds of phenomena need to be represented by different geometries. The notions of geometric and location abstraction are introduced. Geometric and location abstraction are used to relate different geometries and their representations to each other. In the analysis a conceptual model a model level and a representation level are distinguished. The results of the analysis are applied to three classes of problems in GIS: handling of uncertain data and effective organization of spatial data. It is shown that the analysis on different levels provides a unified view on the representation of geometries of geographic space in GIS.

1. INTRODUCTION

Geographic Information Systems (GIS) can be seen as implementations of formal models of geometries of geographic space. In this paper we discuss the application of concepts of geometry to the representation of phenomena in geographic space. This discussion has a long history in spatial data handling and spatial information theory (Frank and Kuhn 1995), e.g., (Corbett 1979), (Frank and Kuhn 1986), (Frank 1987), (Franklin 1990), (Couclelis 1992). We continue this discussion and make different concepts and their relations explicit:

- geometry, geometric abstraction, and representation of geometry;
- location, location abstraction, preserving geometry in terms of location, and representation of location;

The notions of geometric abstraction and location abstraction are introduced in this paper. These notions relate different geometries to each other.

There is no single geometry of geographic space. Different kinds of phenomena are represented by different geometries. We discuss several aspects of the representation of geometries of geographic space:

- What class of phenomena can be represented by a given geometry?
- Which geometries can be represented on a computer and how are they represented?
- What properties do the representations have?

In this discussion we use concepts of geometry, location, and abstraction as well as concepts of theoretical computer science such as finite representability, decidability, and complexity. We separate the discussion on the *conceptual level* using the notions of geometry, location, and abstraction from the discussion on the *representation level* based on the notions of (formal) language, data structures, and effectiveness. Both levels are related to each other by *modeling* concepts of geometry in the domain of point sets and *representing* point sets by means of formal language (modeling level).

We relate the discussion to concrete problems of representation of geographic space in GIS. We consider two problem areas: Handling uncertainty of spatial data and effective organization of spatial data. We demonstrate that in both problem areas the capability of different representations of geometry to *abstract* from aspects which can cause inconsistency or inefficiency is used to achieve a finite and efficient representation.

2. GEOGRAPHY AND GEOMETRIES OF GEOGRAPHIC SPACE

2.1 The Concept of Geometry

Geography considers the location of spatial phenomena on Earth. Location on Earth is measured ,e.g., by surveyors, and transformed into a frame of reference of the plane. Geometry abstracts from particular location and considers properties of *geometric figures* that are independent of particular location, i.e., invariant under certain groups of transformations (Yaglom 1962), (Table 1). Geometric figures are equivalence classes of geometric individuals or configurations of geometric individuals that can be made coincide by transformations belonging to such a group (e.g. triangles, configurations of triangles, polygons, configurations of polygons,…). Geometric properties of configurations, i.e., n-tuples, of geometric figures are usually called spatial relations.

Transformations are operations on geometric individuals that change certain properties and leave others invariant. For example, a translation applied to a configuration of two triangles in the plane changes the location of both triangles but preserves their size and the relative distance of their vertexes. Direct isometries, i.e., rotation and translation, applied to a configuration of triangles do not change the distance of the vertexes but in the case of rotation the order of the vertexes projected on a straight line is changed (Goodman and Pollack 1993), (Schlieder 1995).

A group of transformations is a set of transformations, T, e.g., the set of all translations of triangles in the plane, with a composition operation that combines two transformations to a more complex transformation, $t_3=t_1 \circ t_2, t_i \in T$. The composition operation has the following properties:

- it is closed for the set of transformations, e.g., the composition of arbitrary translations is again a translation;
- it is associative, i.e., the order in which primitive transformations are composed to complex transformations does not matter: $t_1 \circ (t_2 \circ t_3)=(t_1 \circ t_2) \circ t_3$;
- there exists an identity, $0 \in T$: the identity transformation that transforms any configuration of geometric individuals onto itself, e.g., the translation of distance zero;
- every transformation t has an inverse transformation $t^{-1} \in T$, such that the composition of both is the identity transformation: $t \circ t^{-1}=0$.

A geometry is the study of properties of geometric figures, or configurations of geometric figures that remain invariant under a **particular** group of transformations (Klein 1872). There are several groups of transformations which create different geometries (Table 1).

geometry	group of transformations	invariant properties
location-invariant (L)	identity map	location
translation-invariant (T)	translation	distance, order type of point-configurations, congruence, similarity, volume
direct-isometry (D) ‘rigid body movements’	rotation, translation	distance, clockwise order of figure vertexes, volume, congruence , similarity
isometries (I)	rotation, translation, reflection	distance, volume, congruence, similarity
similarities (S)	isometries + zooming	angles, ratio of distance, similarity
affine (A)	linear transformations	coliniarity, neighborhood
isotropies ($\mathfrak{I}$)	homeomorphisms without reflection	neighborhood, embedding in plane, clockwise order of figure vertexes
topology (H)	homeomorphisms	neighborhood, connectedness

Table 1: Different Geometries

GIS are tools to represent and model geographic space on a computer. A geometry is represented as (abstract) data type. (Abstract) data types provide means to represent the invariant properties of the corresponding class of geometric figures implicitly or explicitly and provide means to perform operations, e.g., transformations, on them. The operations are assumed to keep invariant properties unchanged. Geometric individuals are represented as instances of a particular data type.

2.2 Modeling the Geometry of Spatial Phenomena

Representations of space contain knowledge of geometry and location of spatial phenomena and their change. Space, location, geometry, spatial extend, and spatial change is *modeled* in the domain of set of points constituting an *abstract space* and its geometry. Abstract space is constituted by the whole set of points populating it. Spatial change is modeled by transformations of the points constituting the abstract space. Transformations are one-one mappings of the abstract space onto itself.

The mathematical concept of a bijective (one-one and onto) mapping of a set onto itself is a permutation (Piff 1991). The permutations of a set and the composition of permutations form a group, the symmetric group SYM. A property of sets that is preserved under arbitrary permutations is its subset structure. In the remainder geometric transformations are modeled as subgroups of the permutation group. A subgroup of the symmetric group is a group formed by a subset of permutations that leave some properties of point sets invariant and change others. The *structure* of the abstract space is formed by the base-set of points constituting the space, the class of its subsets, and a sub-group of the permutation group of the base-set. The abstract space and its structure are used to represent the geometry of a class of geometric figures.

The spatial extend of geometric figures is modeled by subsets of points which are supposed to fulfill certain properties. The concept of subsets of the set of points of a space is very general. Most of those subsets cannot be used to model the extend of spatial phenomena, e.g., sets of points in the plane which members are arbitrarily scattered. In order to model the extend of spatial phenomena one needs a notion of neighborhood. Point set topology is the study of subsets of a point set using the notion of neighborhood (Kuratowski 1962). Open and closed sets are distinguished in terms of neighborhood. Intuitively, closed sets are sets representing regions with their boundary. Regular closed sets are closed sets which represent regions which boundaries are 'well shaped'. Regular closed sets of a 2D space are used to model regional phenomena because they meet the following intuitions:

- a region has an interior (the points that have no contact to the exterior),
- a region has a boundary (points that have contact to the exterior),
- the boundary encloses the interior, there are no infinitely thin spikes, no single disconnected points.

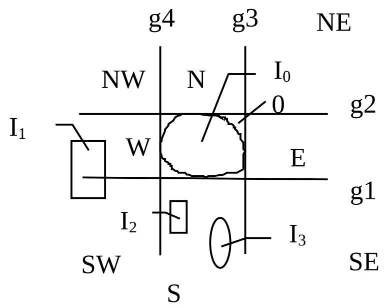
Consider the group of permutations of the points of the plane. In general regular closed sets are not preserved. Consider for example a regular closed set R and a permutation that leaves all points unchanged except two points: the position of two points $p_1 \in R$ and $p_2 \notin R$ are exchanged. Before the permutation $p1$ was located in the interior of R and p_2 in the exterior. After the permutation the set membership of p_1 and p_2 is unchanged but R now has a separated point in its exterior and is not an regular closed set anymore.

The property of being a regular closed set is preserved by the subset of permutations that preserve neighborhood. This class of permutations is called homeomorphisms, i.e., bicontinuous one-one mappings of the points of a space onto itself. The corresponding group is called the group of homomorphisms. The corresponding geometry is the topology of regular closed sets.

2.3 The Concept of Location

Geometry is a technique of abstraction from *individuals* and configurations of individuals to *geometric figures*, i.e., *equivalence classes* of individuals and configurations of individuals with respect to *common, i.e., invariant, properties*. Location is a relation between spatial individuals and a frame of reference. Location within a frame of reference allows to abstract from individuals using equivalence classes with respect to identity of location. Individuals that have the same location in a particular frame of reference form an equivalence class.

Consider Figure 1. A frame of reference is defined by the lines $\{g_1, \ldots, g_4\}$. These lines generate a partition of the plane. The partition regions are labeled: $N, NE, \ldots NW, 0$. $\{I_1, \ldots, I_4\}$ are individuals. The location of the individuals within the frame of reference given by $\{g_1, \ldots, g_4\}$ is defined to be the set of labels of partition regions they intersect. Labels of partition regions are coordinates denoting positions in the frame of reference. Sets of positions occupied by individuals define their location. The domain of location is the powerset of the set of coordinates {N,NE,…NW,0}. Two individuals are equivalent if they have the same location (e.g., I_2 and I_3). In this example the domain of location is finite. It has the size of the powerset of the set of coordinates. The infinite domain of individuals ($\{I_0,\ldots,I_4\}$ are only examples) is subdivided into a finite set of equivalence classes with respect to location within the frame of reference given by $\{g_1,\ldots,g_4\}$.



$LOC(I_0)=(0)$
$LOC(I_2)=LOC(I_3)=(S)$
$LOC(I_1)=(W,SW)$

Figure 1: Location of Individuals within a Frame of Reference (e.g., [Frank, 1992 #36])

There are different types of frames of reference. Often frames of reference define a finite partition of space. The semantic of the coordinates denoting positions depends on the embedding of the frame of reference into space, the structure of the frame of reference, and the definition of positions. Consider Figure 1. Assuming that the lines g_1 and g_2 are aligned to projections of the latitude and g_3 and g_4 are aligned to projections of the longitude then the location N can be interpreted as North with respect to the location 0 (interpreted as location of the reference object).

An important issue is to evaluate which geometric properties of the domain of individuals are preserved in the domain of equivalence classes of individuals with respect to location, i.e., in the domain of location. The degree until which the geometry of the domain of individuals is preserved depends on the structure of the coordinate space, i.e., the frame of reference and the definition of location. Consider Figure 1: For geometric individuals that share a location no geometric distinctions are preserved at all. Nothing can be said about geometric properties of the configuration (I_2,I_3), neither about distance nor about topological relations.

An important special case is location within a frame of reference given by a system of n directed axes of real numbers. The coordinate space created by such a frame of reference is $\mathbf{R}^n$, i.e., the n-ary cross product of real numbers. Locations are sets of n-tuples of real numbers. An important property is that the coordinate space is 'fine ' enough to distinguish arbitrary geometric individuals that can be modeled by means of semi-linear sets (see below). The equivalence classes with respect to location contain exactly one element. The complete geometry of the domain of individuals is preserved in the domain of location.

3. Representation of Geometry

Representing geometric figures, modeled as point sets, on a computer means to represent possibly infinite sets of points by finite means of language. The relations between language representations and their corresponding structures are studied in model theory (Hodges 1997). The model theory of finite representable domains relevant in the context of representation of geometry is reviewed in (Grumbach and Su 1994).

In this section different representations of geometry are discussed. In order to evaluate different *geometric* representation and reasoning methods a number of specific questions need to be analyzed (Papadimitriou, Suciu et al. 1996):

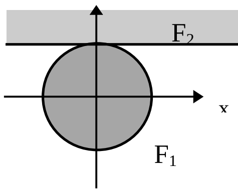
- Which geometry is formalized and how complete are the properties of this geometry represented, i.e., what class of figures are representable and which properties of these figures remain invariant? A formalism (data structure) represents a geometry *completely* if *arbitrary* geometric configurations represented by the same formula (data instance) can be transformed into each other by means of computable transformations belonging to the group of transformations defining the particular geometry.

- If the representation of the geometry is incomplete - is it complete for a significant class of geometric figures or configurations of geometric figures? Can a subclass of figures, e.g., rectangles, in this geometry be represented completely or only certain configurations, e.g., pairs or triples of figures?
- Which properties of semi-algebraic sets (see below) can be completely specified using the representation discussed? Semi-algebraic sets play a special role in geometric representation and modeling. They are the largest class of point sets (1) which geometry of direct isometries can be *defined finitely and completely* in terms of location in $\mathbf{R}^n$, and (2) for which exists an *automatic* reasoning procedure (Kanellakis, Kuper et al. 1992).

3.1 Representation of Geometry in Terms of Location in $\mathbf{R}^n$

Representing location of geometric figures on a computer quantitatively means: To represent possibly infinite sets of points, modeling geometric figures, in terms of their location within an ordering frame of reference provided by axes of real numbers, by finite means of language. Analytical geometry is a mathematical theory of geometric properties of sets of points based on finite representations of their locations in the frame of reference given by a system of n coordinate axes of directed real lines and the corresponding coordinate space $\mathbf{R}^n$. In general analytical geometry cannot be completely implemented on a finite computer since real numbers cannot be represented finitely (Franklin 1984). Recently a number of sub-languages of analytical geometry were investigated which can be implemented on a computer effectively (Kanellakis, Kuper et al. 1992). Sublanguage means that the expressive power of the language is constrained to represent a smaller class of geometric figures and to operations that do not rely on the instantiation of real number values.

There are approaches, e.g., (Kuijpers, Paredaens et al. 1995), that consider as a geometrical figure any figure that is definable in elementary geometry, i.e., first order logic over the real numbers. The corresponding class of figures is the class of semi-algebraic sets. Semi-algebraic sets (Bochnak, Coste et al. 1987) have the form $F = \{ (x_1,\ldots,x_m) \mid x_1,\ldots,x_n \in \mathbf{R}, \Phi(x_1,\ldots,x_m) \}$. The formula Φ *defines* the set F. The members of the set F are all n-tuples $(x_1,\ldots,x_n)$ that satisfy the formula Φ, i.e., make it true. In this paper we consider formula Φ which consist of a finite disjunction of conjunctions polynomial inequalities. Polynomial inequalities have the form $P \leq Q$ where P and Q are polynomials in real variables with rational coefficients (Figure 2). Formula Φ of this structure are called semi-algebraic formula. In the remainder of this paper these formula are called FO+poly. The important point is that the coefficients (numbers explicitly occurring in the formula) need to be rational numbers that can be represented finitely (Franklin 1984). Operations on geometric figures, modeled by point sets, represented by formula are defined in terms of operations on formula. Often operations can be defined *without explicitly* computing coordinate values.



$F_1 = \{ (x,y) \mid x^2 + y^2 \leq 1 \}$
$F_2 = \{ (x,y) \mid 0 \leq y - 1 \}$

Figure 2: Semi-Algebraic Sets

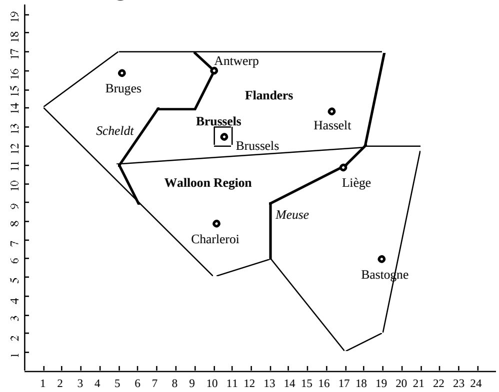
Semi-algebraic sets are the largest subclass of regular closed sets which are finitely representable by formula. The important point is that the *finite* representability of infinite sets is due to the structure of the domain of location, i.e., the algebraic structure of $\mathbf{R}^n$. Semi-algebraic sets are defined, i.e., represented, finitely and implicitly in terms of formula representing the *location* of its members in a frame of reference. The frame of reference is defined in terms of a system of n coordinate axes of directed real lines. Semi-algebraic sets are represented in the *domain of location* (pairs of real numbers) rather than in the domain of individuals (points of the plane).

Reasoning about semi-algebraic sets in terms of operations on formula FO+poly is decidable (Tarski 1951), i.e., can performed automaitcally, but has a very high complexity. There are a number of sub languages with lower complexity which were studied for example in (Afrati, Cosmadakis et al. 1994), (Vandeurzen, Gyssens et al. 1995). The most interesting sublanguage for practical applications is called FO+linear. In FO+linear polynomials are linear and the coefficients range over rational numbers. The corresponding point sets are semi-linear sets (e.g., F_2 in Figure 2). It was shown that for semi-linear sets represented by linear polynomial inequalities, i.e., linear constraints, effective reasoning procedures exist (Afrati, Cosmadakis et al. 1994).

Linear inequalities represent halfplanes. Conjunctions of linear inequalities represent intersections of halfplanes. Finite conjunctions of linear inequalities represent convex polygons. Disjunction of conjunctions of linear inequalities represent non-convex polygons (Guenther 1987), (Vandeurzen, Gyssens et al. 1995). An example is given in Figure 3. The advantage of linear constraint representations (FO+linear) is that they can be represented on a computer correctly and completely. Operations such as union, intersection, buffering, … of semi-linear sets can be performed effectively without approximation (Haunold, Grumbach et al. 1997). The approximation of real numbers can be postponed until particular real coordinates need to be instanciated, for example, in order to compute the area of a region or to draw boundaries.

A disadvantage is that the representation is highly implicit: From the formal representation in Figure 3 it is difficult to see which formula defines which halfplane. For example, a boundary segment shared by two regions needs to be computed by intersection of halfplanes before it can be drawn. More abstractly: There is *no explicit representation of geometric invariances*. This has the consequence that operations involving invariants cannot be applied to parts of the representation but must be computed every time it is used, e.g., neighborhood of polygons sharing a common edge needs to be computed from the intersection of halfplanes. From this point of view representations based on linear constraints (FO+linear) have some of the disadvantages of spaghetti representations (Laurini and Thompson 1994).

The implicit character of the representation is due to the fact that the representation is totally committed to the domain of location. The domain of location does not ‘know’ about the geometry of individuals. *It preserves it implicitly* in the structure of the formula and the values of the parameters of the polynomials. Knowledge of the geometry of individuals can be derived from knowledge of their location, for example we can *compute* whether the location `(x,y)` is the solution of polynomials defining different semi-linear sets and *conclude* that the corresponding geometric figures are connected.



Name	Geometry
Brussels	$(y \le 13) \wedge (x \le 11) \wedge (y \ge 12) \wedge (x \ge 10)$
Flandres	$(y \le 17) \wedge (3x - 4y \ge -53) \wedge (x - 14y \le -150) \wedge (x + y \ge 45) \wedge$ $(4x - y \le 78) \wedge (\neg((y \le 13) \wedge (x \le 11) \wedge (y \ge 12) \wedge (x \ge 10)))$
Walloon Region	...

Figure 3: Representation of regions using inequalities (Vandeurzen, Gyssens et al. 1995)

3.2 Representation of Geometry Abstracting from Location in $\mathbf{R}^n$

Semi-algebraic sets can be explicitly *defined* (and implicitly finitely represented) using formula of the class FO+poly. This is due to the algebraic structure of the coordinate space $\mathbf{R}^n$ on which these techniques heavily depend on. The commitment to the algebraic structure of the coordinate space $\mathbf{R}^n$ on the level of representation implies the commitment to the concepts of location on the conceptual level. Formally location is defined within the frame of reference provided by a system of n directed real axes.

The commitment to the representation in terms of FO+poly (FO+linear and equivalent representations) constraints the class of sets representable to the small class representable with semi-algebraic (semi-linear) sets. Suppose the regional phenomena in geographic space are modeled using regular closed sets then only the small class of semi-algebraic (semi-linear) sets can be represented using the techniques discussed in section 3.1. There is a large class of regular closed sets, and corresponding geographic phenomena, that either need to be approximated or cannot be represented at all (see section 4.1 for further discussion).

If we give up the concept of location within $\mathbf{R}^n$ then we cannot use its algebraic structure for representation and reasoning. For a discussion why to abstract from location in $\mathbf{R}^n$ see section 4. We continue the discussion in this section on the conceptional and the representation level. Abstraction from semi-algebraic sets and their location in $\mathbf{R}^n$ is possible in the domain of location or in the domain of individuals. Abstraction in the domain of individuals means to chose another geometry, i.e., make less distinctions between individuals, represent a larger class of geometric figures. Abstraction in the domain of location means to chose another frame of reference and another definition of location.

3.2.1 Abstraction in the Domain of Individuals

Geometric abstraction from geometry geo_1 to geo_2 means to represent information of geometrical figures of a geometry geo_1 with transformations t_i belonging to a group of transformations $Gr_1 \subseteq SYM$, $t_i \in Gr_1$, in another geometry geo_2 with transformations $t_j, t_j \in Gr_2$, $Gr_2 \subseteq SYM$, such that Gr_1 is a *subgroup* of Gr_2, i.e., $Gr_1 \subseteq Gr_2$. Geometric representations are supposed to be *complete* for *arbitrary* n-tuples, i.e., configurations of geometric individuals in this class. The domain of geometric individuals is divided into equivalence classes of configurations that can be transformed into each other by means of a group of operations defining the particular geometry.

Representations of geometries abstracting from semi-algebraic sets and their location in $\mathbf{R}^n$ are based on *symbolic structures that explicitly represent invariant properties* of geometric figures: Symbolic structures act as invariants explicitly characterizing equivalence classes of a class of sets with respect to a group of transformations[1]. This means that the transformations applied to a geometric configuration do not change its symbolic representation.

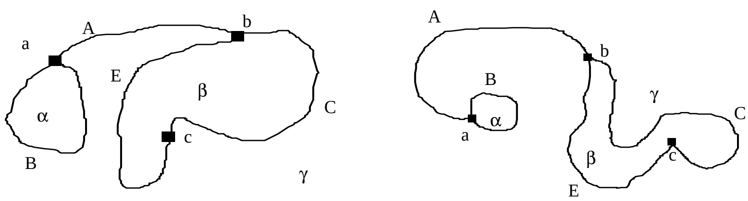


Figure 4: Two equivalent PLA-structures (Hidders 1995)

Consider Figure 4: There are two graph structures representing the geometry of two different configurations of spatial individuals. Spatial individuals are modeled by subsets of points of the plane: α, β, γ refer to two dimensional regular open sets. A, B, C, E refer to one dimensional open sets. a, b, c refer to points of the plane. α, β, γ, A, B, C, D, a, b, c are symbols *explicitly* naming points and point sets in the plane. A spatial configuration is explicitly represented by an unidirected multi-graph with: Edges labeled by names of one dimensional open sets, nodes labeled by names of points where one dimensional sets intersect, begin, or end, and faces labeled with names of two dimensional open sets enclosed by edges.

The graph structure is a symbolic structure explicitly representing geometric properties that remain invariant under isotropic transformations (topological transformations of the plane). Both configurations in

1 The original formulation considering topological invariances can be found in (Papadimitriou, Suciu et al. 1996, pg. 82)

Figure 4 are represented by the same graph. Both configurations can be transformed into each other by means of planar topological transformations. The graph representation is not complete for isotropic transformations (Hidders 1995), i.e., there are configurations that are represented by the same graph structure but cannot be transformed into each other by means of isotropic transformations. PLA-structures (Kuijpers, Paredaens et al. 1995) are extended graph representations explicitly representing for every point the cyclic order of its neighboring edges and faces (for point `a` in Figure 4: `[α,B,γ,A,γ,B]`) as well as the name of the one non-bounded region (γ in Figure 4). PLA structures are complete for the geometry of isotropies (Hidders 1995). They represent isotropies of regular closed sets (of dimension 0, 1, 2) completely and explicitly. They are more abstract than representations of semi-algebraic sets since they represent a larger class of sets and distinguish less properties. PLA-like structures are widely used in GIS, e.g., (Corbett 1979), (Herring 1987), (Morehouse 1989), (Frank and Kuhn 1986).

3.2.2 Abstraction in the Domain of Location

Abstraction in the domain of location (*location* abstraction) means to chose a frame of reference and define location such that in the domain of location a more *general* geometry for a *fixed set of individuals* is preserved. The critical point of abstraction in the domain of location is to chose the 'appropriate' frame of reference. Appropriateness is evaluated with respect to the representation the geometry of a finite set of individuals. A frame of reference is 'appropriate' to represent the geometry of a set of individuals if the domain of location is sufficient

- to distinguish all the selected individuals,
- to preserve the intended geometry for the set of individuals it is supposed to distinguish.

Consider Figure 1: the domain of location is not sufficient to distinguish between `I2` and `I3`. Ignore `I3`. Then the domain of location preserves the topological properties of the following set, `C`, of configurations: `C = {(I0,I1),(I1,I2),(I0,I2)}`. From the structure of the frame of reference, the definition of the location, and the knowledge of the their location, the disjoint-relation can be derived for the configurations in `C`. The frame of reference is sufficient to preserve cardinal directions, i.e., translation-invariant geometric properties. From knowledge of location and from knowledge of the embedding of the frame of reference, i.e., the alignment of the axes g1,…,g4 to latitude and longitude cardinal directions can be derived for the set of configurations `C`.

approach	regional partition	domain of location
(Winter 1995)	regular raster + equidistance	$Loc \subset Pow(Z^2)$
(Papadias and Sellis 1993)	regular raster adjusted to latitude and longitude	$Loc \subset Pow(Z^2 \times \{\delta, °\})$
[Frank, 1992 #36]	regular 3×3 raster adjusted to latitude and longitude	Pow({N,…,NW,0})
(Freksa 1992)	regular 2×3 raster adjusted to vector between two points	Pow({left,right}×{front,same,back})
(Hernandez 1994)	cone shaped raster	Pow({back,right-back,…,left-back}× {containment,overlap,tangent,disjoint})

Table 2: Examples of Abstraction in the Domain of Location

Consider location in the plane. In general abstraction from location in a frame of reference given by a system of 2 axes of real numbers leads to location in frame of reference formed by a finite regional partition of the plane. Regional partitions are not necessarily regular, i.e., partition elements can be arbitrarily shaped regions. A special instance is the regular partition formed by 9 regions in Figure 1. A general framework of representing location of regional individuals within the frame of reference given by the regional partition was given in (Bittner 1997), (Bittner 1997a). Examples of applications of location based abstraction for special domains of location are given in (Table 2).

4. GEOMETRIC ABSTRACTION IN GIS

In this section we discuss how techniques of geometric abstraction and location abstraction,are used in GIS to abstract from geometric aspects of the representation of spatial phenomena which are unknown or indeterminate (section 4.1), add distinctions that are not relevant for the task in hand (section 4.2), unnecessarily increase the complexity of the representation (section 4.2).

4.1 Uncertainty of Spatial Data

The special nature of the phenomena represented in GIS is that they might have undetermined boundaries or do not have any boundaries at all (Burrough and Frank 1996). Spatial phenomena with undetermined boundaries can be partitioned into three broad categories: Objects which do have a sharp boundary but which positions and shape are unknown or cannot be measured exactly [positional uncertainty]; Objects which do not have a well defined boundary or for which it is useless to fix a boundary [definition uncertainty] (Clementini and Di Felice 1996, pg. 2); Phenomena without boundaries, i.e., continuously distributed spatial phenomena, e.g., temperature, humidity of the air, etc.

4.1.1 Geometric Representation of Positional Uncertainty

Positional uncertainty can be handled by means of geometric abstraction or location abstraction. Resulting geometries are those which properties remain invariant under translations. They should be complete since the geometric configurations themselves are assumed to be well defined (the definition is certain).

Geometric abstraction, i.e., by representing a location invariant rather a location depended geometry. The degree of abstraction can vary from translation-invariant geometry (abstracting from location and preserving relative distance, shape, volume, …) to isotrophy-invariant geometries (preserving the topology of a spatial configuration and their embedding in the plane). From measurement of location is *abstracted totally*. Representation of geometric properties is independent of measurement of location.

Location abstraction, i.e., by abstracting from location in an Euclidean frame of reference; The degree of abstraction can vary from location in a more or less fine regular raster, e.g., (Winter 1995), to location in a non regular partition. An example for location in a non-regular partition is location in a political subdivision (‘the area of pollution is in Idaho’) or classification of soil-types. *Measurement of location* is replaced by *inspection of occupation* of (parts of) partition elements (Bittner 1997).

4.1.2 Definitional Uncertainty

Definitional uncertainty concerns with phenomena not having a well defined boundary. The phenomena and their boundaries cannot be modeled by the topology of regular closed sets. The main reason is that there is no concept of degree of membership in a set. A point either belongs to a set modeling the extend of a valley or not. Examples are valleys, mountains, areas of pollution, the Atlantic Ocean., etc. There are several possibilities to handle these problems:

- Extend the concept of set membership, e.g., fuzzy sets (Fisher 1996). The geometry of fuzzy sets is not well understood yet.
- Approximate the phenomena by regular closed sets and chose a geometric representation that makes the least commitments, i.e., representations of the topology of regular closed sets (minimize explicitly wrong representations).
- Approximate spatial phenomena by configurations of geometric figures representing the geometry of a ‘certain’ core and an extended ‘uncertain boundary’. These geometries of are again subject to positional uncertainty.

4.1.3 Phenomena without Boundaries

Phenomena that do not have boundaries are quite common in geographic space. Examples are temperature, soil type, soil wetness. Geographers model these kinds of phenomena as fields. Fields, i.e., continuous distributed phenomena, are modeled using continuous functions. Due to the finite and discrete character of computer representations only approximations of these models can be implemented. Finite approximations are for example polynomials with rational coefficients or regional partitions. The regional partition is either created by a frame of reference resulting in regular partitions (e.g., raster) or by subdividing the target domain (e.g., domain of temperature) into equivalence classes resulting in non-regular regional partitions. Finite geometric representations are PLA-model-like graph structures representing regional partitions of the plane, e.g., cell complexes.

Relations between phenomena with and without boundaries can be represented using the concept of location: Continuously distributed phenomena are modeled by a regional partition of space. Regional partitions provide frames of reference in which spatial phenomena with determinate or indeterminate boundaries are located. The *relation* between a bounded and non-bounded phenomenon is the *location* of the bounded phenomenon within the frame of reference given by the regional partition approximation of the non-bounded phenomenon (Bittner 1996)

4.2 Organization of Spatial Data

Effective organization is critical for the management of large amount of spatial data (Frank 1984). Spatial indexes can be considered as location based representations of geometry of abstract geometric figures. Consider R-trees (Guttman 1984). R-trees are based on rectangular partitions of the plane. Regional partitions form a frame of reference. R-trees represent location in regional partitions. Positions are names of partition elements. Location of an individual is defined in terms of containment in a partition region. Location is represented by single positions. The special feature of the domain of location is that it is dynamic. The structure of the partition may change. This causes changes in the domain of location and the corresponding R-tree representation. Changes of the partition structure are caused by violation of the following constraints:

- For every individual represented a location must be defined. Location is defined in terms of *containment* in partition regions. If an individual to be inserted *intersects* boundaries the partition structure the partition is reorganized.
- Every location is occupied by at most n individuals. If the number is exceeded the partition is refined.

If one of the constraints is violated the partition structure is reorganized. This causes a reorganization of the domain of location and the R-tree structure.

Consider the domain of individuals. Individuals are rectangles in the plane. Rectangles are the result of geometric abstraction. Each rectangle is the equivalence class of all individuals that have the same minimum-bounding-rectangle (MBR), i.e., the set of all individuals which MBRs coincide.

Consider location of blocks on the disk. The set of blocks form a regional partition of the disc-space consisting of n circles. They form a frame of reference. Positions are blocks. The coordinate space is given by the block numbers. Location is defined by sets of positions occupied. Neighborhood in the domain of location, i.e., positions of blocks on the disk, is assumed in the domain of location given by sets of natural numbers.

Effective disk access is achieved by *correspondence of the two domains of location*: the domain of location of blocks on the disk and the domain of location of MBRs in a regional partition of the plane represented in a R-tree. Correspondence means that there exists a one-one mapping between the locations in both domains. The better this mapping preserves neighborhood in the domain location in the R-tree domain in the domain of location on disk the more effective is the access (Frank 1984).

Spatial indexing methods are the result of geometric abstraction in the domain of geometric individuals , resulting in MBRs. Location abstraction in the domain of MBRs results in R-tree representations of location of rectangles in a frame of reference formed by a rectangular partition of the plane. The existence of a mapping between the resulting domain of location of MBRs and a domain of location of blocks on disk supports fast disc access and supports effective handling of spatial data.

5. SUMMARY

In this paper we argued that different classes of phenomena need to be represented by different geometries. We discussed techniques of geometric and location abstraction and their application to the representation of spatial phenomena of geographic space in GIS. These techniques provide means to relate different geometries and their representation to each other. We discussed the application of these concepts to problems of handling uncertainty of spatial data and effective organization of spatial data.

6. REFERENCES

Afrati, F., S. S. Cosmadakis, et al. (1994). Linear vs. Polynomial Constraints in Database Query Languages. PPCP 1994.

Bittner, T. (1996). A Qualitative Model of Geographic Space. 7th International Symposium on Spatial Data Handling, Delft, The Netherlands.

Bittner, T. (1997). A Qualitative Coordinate Language of Location of Figures within the Ground. Spatial Information Theory, Springer.

Bittner, T. (1997a). Towards a Model Theory for Figure Ground Location. 6th Symposium on Mathematics and AI, Forth Lauderdale, Florida.

Bochnak, J., M. Coste, et al. (1987). Géométrie algébrique réelle. Berlin ; New York, Springer-Verlag.

Burrough, P. and A. U. Frank, Eds. (1996). Geographic Objects with Indeterminate Boundaries. GISDATA, Taylor &Francis.

Clementini, E. and P. Di Felice (1996). An algebraic model for spatial objects with undetermined boundaries. Geographic Objects with Indeterminate Boundaries. P. Burrough and A. U. Frank, Taylor & Francis. **2**.

Corbett, J. P. (1979). Topological principles of cartography. Washington, SC, USA, US Bureau of the Census.

Couclelis, H. (1992). People Manipulate Objects (but Cultivate Fields): Beyond the Raster-Vector Debate in GIS. Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Springer.

Fisher, P. (1996). Boolean and Fuzzy Regions. Geographic Objects with Indeterminate Boundaries. P. Burrough and A. U. Frank, Taylor & Francis. **2**.

Frank, A. U. (1984). Requirements for Database Systems Suitable to Manage Large Databases.

Frank, A. U. (1987). Overlay Processing in Spatial Information Systems. Eighth International Symposium on Computer-Assisted Cartography (AUTO-CARTO 8), Baltimore, MD.

Frank, A. U. and W. Kuhn (1986). Cell Graphs: A Provable Correct Method for the Storage of Geometry. Second International Symposium on Spatial Data Handling, Seattle, WA, IGU.

Frank, A. U. and W. Kuhn, Eds. (1995). Spatial Information Theory - A Theoretical Basis for GIS. Lecture Notes in Computer Science, Springer.

Franklin, R. (1990). Calculating Map Overlay Polygons' Areas without Explicitly Calculating the Polygons - Implementation. SDH, Zuerich, Swizerland.

Franklin, W. R. (1984). Cartographic Errors Symptomatic of Underlying Algebra Problems. First International Symposium on Spatial Data Handling, Zurich, Switzerland.

Freksa, C. (1992). Using Orientation Information for Qualitative Spatial Reasoning. Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. A. U. Frank, I. Campari and U. Formentini, Springer.

Goodman, J. E. and R. Pollack (1993). Allowable Sequences and Order Types in Discrete and Computational Geometry. New Trends in Discrete and Computational Geometry. J. Pach, Springer.

Grumbach, S. and J. Su (1994). Finitely Representable Databases. ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems.

Guenther, O. (1987). Efficient Structures for Geometric Data Management, Springer Verlag.

Guttman, A. (1984). “R-Trees: A Dynamic Index Structure for Spatial Searching.” SIGMOD'84 **14**(2).

Haunold, P., S. Grumbach, et al. (1997). Linear Constraints: Geometric Objects Represented by Inequalities. Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '97, Laurel Highlands, PA. S. Hirtle and A. Frank. Berlin, Springer-Verlag. **1329:** 429-440.

Hernandez, D. (1994). Qualitative Spatial Reasoning, Springer.

Herring, J. R. (1987). TIGRIS: Topologically Integrated Geographic Information System. Auto-Carto 8, Baltimore, Maryland.

Hidders, J. (1995). An Isotropic Invariant for Planar Drawings of Connected Planar Graphs, Eindhoven University of Technology.

Hodges, W. (1997). A Shorter Model Theory. Cambridge, Cambridge University Press.

Kanellakis, P. C., G. M. Kuper, et al. (1992). “Constraint Query Languages.” .

Klein, F. (1872). Vergleichende Betrachtungen über neuere geometrische Forschungen. Erlangen, Germany.

Kuijpers, B., J. Paredaens, et al. (1995). Lossless representation of topological spatial data. Advances in Spatial Databases, Springer.

Kuijpers, B., J. Paredaens, et al. (1995). Semantics in Spatial Databases. Semantics in Databases. B. Thalheim, Springer.

Kuratowski, K. (1962). Introduction to Set Theory and Topology. Oxford, Pergamon Press LTD.

Laurini, R. and D. Thompson (1994). Fundamentals of Spatial Information Systems, Academic Press.

Morehouse, S. (1989). The Architecture of ARC/INFO. AUTO-CARTO 0, Baltimore, Maryland.

Papadias, D. and T. Sellis (1993). The semantics of relations in 2D space using representative points : spatial indexes. Spatial Information Theory, A Theoretical Basis for GIS, Elba, Springer.

Papadimitriou, C. H., D. Suciu, et al. (1996). Topological Queries in Spatial Databases. PODS'96, Montreal, ACM.

Piff, M. (1991). Discrete Mathematics, Cambridge University Press.

Schlieder, C. (1995). Reasoning About Ordering. Spatial Information Theory - A Theoretical basis for GIS, Semmering, Austria, Springer.

Tarski, A. (1951). A Decision Method for Elementary Algebra and Geometry.

Vandeurzen, L., M. Gyssens, et al. (1995). On the Desirability and Limitations of Linear Spatial Database Models. Fourth International Symposium on Large Spatial Databases - SSD95, Portland, ME, Springer Verlag, Heidelberg.

Winter, S. (1995). Topological Relations between Discrete Regions. Advances in Spatial Databases, SSD'95, Portland, ME, USA.

Yaglom, I. M. (1962). Geometric Transformations, Random House.