Uncertainty Management for Spatial Data in Databases: Fuzzy Spatial Data Types

Markus Schneider

FernUniversität Hagen, Praktische Informatik IV
D-58084 Hagen, Germany
Markus.Schneider@fernuni-hagen.de

Abstract. In many geographical applications there is a need to model spatial phenomena not simply by sharply bounded objects but rather through vague concepts due to indeterminate boundaries. Spatial database systems and geographical information systems are currently not able to deal with this kind of data. In order to support these applications, for an important kind of vagueness called *fuzziness*, we propose an abstract, conceptual model of so-called *fuzzy spatial data types* (i.e., a *fuzzy spatial algebra*) introducing *fuzzy points*, *fuzzy lines*, and *fuzzy regions*. This paper* focuses on defining their structure and semantics. The formal framework is based on fuzzy set theory and fuzzy topology.

1 Introduction

Representing, storing, quering, and manipulating spatial information is important for many non-standard database applications. Specialized systems like geographical information systems (GIS) and spatial database systems to a certain extent provide the needed technology to support these applications. So far, spatial data modeling has implicitly assumed that the extent and hence the borders of spatial phenomena are precisely determined, homogeneous, and universally recognized. From this perspective, spatial phenomena are typically represented by sharply described *points* (with exactly known coordinates), *lines* (linking a series of exactly known points), and *regions* (bounded by exactly defined lines which are called *boundaries*). Special data types called *spatial data types* (see [Sch97] for a survey) have been designed for modeling these spatial data. We speak of *spatial objects* as instances of these data types. The properties of the space at the points, along the lines, or within the regions are given by attributes whose values are assumed to be constant over the total extent of the objects. Well known examples are especially man-made spatial objects representing engineered artifacts like highways, houses, or bridges and some predominantly immaterial spatial objects exerting social control like countries, districts, and land parcels

* This research was partially supported by the CHOROCHRONOS project, funded by the EU under the Training and Mobility of Researchers Programme, contract no. ERB FMRX-CT96-0056.

with their political, administrative, and cadastral boundaries. We will denote this kind of entities as *crisp* or *determinate* spatial objects.

Increasingly, researchers are beginning to realize that the current mapping of spatial phenomena of the real world to exclusively crisp spatial objects is an insufficient abstraction process for many spatial applications and that the feature of *spatial vagueness* or *spatial indeterminacy* is inherent to many geographic data [BF96]. Moreover, there is a general consensus that applications based on this kind of indeterminate spatial data are not covered by current GIS and spatial database systems. In this paper we focus on a special kind of spatial vagueness called *fuzziness*. Fuzziness captures the property of many spatial objects in reality which do not have sharp boundaries or whose boundaries cannot be precisely determined. Examples are natural, social, or cultural phenomena like land features with continuously changing properties (such as population density, soil quality, vegetation, pollution, temperature, air pressure), oceans, deserts, English speaking areas, or mountains and valleys. The transition between a valley and a mountain usually cannot be exactly ascertained so that the two spatial objects "valley" and "mountain" cannot be precisely separated and defined in a crisp way. We will designate this kind of entities as *fuzzy* spatial objects.

The goal of this paper is to present a formal object model for *fuzzy points*, *fuzzy lines*, and *fuzzy regions* in two-dimensional Euclidean space, an effort which is to lead to a *fuzzy spatial algebra*. We propose *fuzzy set theory* and *fuzzy topology* as appropriate conceptual tools for modeling indeterminate spatial data. Fuzzy set theory is an extension and generalization of classical set theory; the approach of fuzzy sets replaces the crisp boundary of a classical set by a gradual transition zone and permits partial and multiple set membership. For fuzzy regions, different views give a better understanding of their nature and also demonstrate how these objects can be represented as (collections of) crisp regions. Consequently, the current exact object models for crisp spatial objects can be considered as simplified special cases of a richer class of models for general spatial objects. It turns out that this is exactly the case for the model to be presented.

Section 2 explains different aspects of spatial vagueness and presents related work. Section 3 introduces some basic definitions of fuzzy set theory and fuzzy topology as far as they are needed in this paper. Sections 4, 5 and 6 formally define fuzzy points, fuzzy lines, and fuzzy regions, respectively. Since the definition for fuzzy regions does not expose their geometric structure, Section 7 provides several structured views of fuzzy regions based on collections of crisp regions. Section 8 draws some conclusions and gives a prospect of future research.

2 Aspects of Spatial Vagueness and Related Work

In current spatial data modeling, the entity-oriented view of spatial phenomena, which we will take in this paper, considers determinate spatial objects as conceptual and mathematical abstractions of real-world entities which can be identified and distinguished from the rest of space. For example, a crisp region partitions space into an interior, a boundary, and an exterior part which are mutually exclusive and cover the whole space. Hence, the notion of a crisp region is intrinsically related to the notion of a boundary. This view fits very well with the mathematical concepts given by the Jordan Curve Theorem and ordinary point set topology.

Boundaries are considered as sharp lines that represent abrupt changes of spatial phenomena and that describe and thereby distinguish regions with different characteristic features. The assumption of crisp boundaries harmonizes very well with the internal representation and processing of spatial objects in a computer which requires precise and unique internal structures. Hence, in the past, there has been a strong tendency to force reality into crisp objects. In practice, however, there is no apparent reason for the whole boundary of a region to be determined. There are a lot of geographical application examples illustrating that the boundaries of spatial objects can be partially or totally indeterminate or blurred. For instance, boundaries of geological, soil, and vegetation units [Alt94, Bur96, KV91, LAB96] are often sharp in some places and vague in others; many human concepts like "the Indian Ocean" are implicitly vague.

In the real world, there are essentially two categories of indeterminate boundaries: sharp boundaries whose position and shape are unknown or cannot be measured precisely, and boundaries which are not well-defined or which are useless (e.g., between a mountain and a valley) and where essentially the topological relationship between spatial objects is of interest. According to these two categories, mainly two kinds of spatial vagueness can be identified: uncertainty and fuzziness. *Uncertainty* is traditionally equated with randomness and chance occurrence and relates either to a lack of knowledge about the position and shape of an object with an existing, real boundary (*positional* uncertainty) or to the inability of measuring such an object precisely (*measurement* uncertainty). *Fuzziness* is an intrinsic feature of an object itself and describes the vagueness of an object which certainly has an extent but which inherently cannot or does not have a precisely definable boundary.

The subject of modeling spatial vagueness has so far been predominantly treated by geographers but rather neglected by computer scientists. At least three alternatives are proposed as general design methods: (1) *exact models* [CF96, CG96, ES97b, Sch96] which transfer type systems and concepts for spatial objects with sharp boundaries to objects with unclear boundaries and which model both uncertainty and fuzziness but in a restricted way, (2) *probabilistic models* [Bla84, Bur96, Fin93, Shi93] which are based on probability theory and predominantly model positional and measurement uncertainty, and (3) *fuzzy models* [Alt94, Bur96, Dut89, Dut91, KV91, LAB96, Use96, Wan94, WHS90] which are all based on fuzzy set theory and predominantly model fuzziness.

The exact object model approach profits from existing definitions, techniques, data structures, algorithms, etc. which need not be redeveloped but only modified and extended. Except for [ES97b], the approaches are based on some kind of *zone* concept. Vague boundaries of a region are modeled as zones expressing the minimal and maximal possible extent of a region. *Vague regions* [ES97b] are a generalization of these models. A vague region is defined as a pair of disjoint, crisp regions. The first region called the *kernel* describes the area which definitely and always belongs to the vague region. The second region called the *boundary* describes the *area* for which we cannot say with any certainty whether it or parts of it belong to the vague region or not. *Maybe* it is the case, *maybe* it is not. Or we could say that this is *unknown*. Vague regions are based on a *three-valued logic*, and boundaries need not necessarily be one-dimensional structures but can be regions.

Probability theory is able to represent uncertainty and defines the membership grade of an entity in a set by a statistically defined probability function. It deals with the *expectation* of a future event, based on something known now. Examples are the uncertainty about the spatial extent of regions defined by some property such as temperature, or the water level of a lake.

Fuzzy set theory deals only with fuzziness. It describes the *admission of the possibility* (given by a so-called *membership function*) that an individual is a member of a set or that a given statement is true. Hence, the vagueness represented by fuzziness is not the uncertainty of expectation. It is the vagueness resulting from the imprecision of meaning of a concept. Examples of fuzzy spatial objects include mountains, valleys, biotopes, oceans, and many other geographic features which cannot be rigorously bounded by a sharp line.

Another difference between fuzzy set theory and probability theory is that in the first case the possibility that an individual belongs to a set depends on *subjective* factors (e.g., expert knowledge) whereas in the second case probability can be computed formally or determined empirically and is thus more *objective*. Moreover, fuzzy set theory enables vague statements about one concrete object whereas probability theory makes statements about a collection of objects from which one is selected. Hence, fuzzy set theory models *local vagueness* while probability theory models *global* vagueness.

The only proposal of a fuzzy data type relates to fuzzy regions [Alt94] defined as a fuzzy set over $\mathbb{N}^2$. Each coordinate $(x, y) \in \mathbb{N}^2$ is associated with a value between 0 and 1 and describes the concentration of some feature attribute at that point. Unfortunately, the simple set property is insufficient since geometric anomalies can arise, as we will see later. The possible importance of fuzzy sets for geographical applications is demonstrated in [Bur96, LAB96, Use96] where also examples of application-specific membership functions are given. The benefits of fuzzy set theory for approximate spatial reasoning and fuzzy query languages is shown in [Dut89, Dut91, KV91, Wan94]. [WHS90] models fuzzy objects by means of the relational data model.

3 Fuzzy Sets and Fuzzy Topology

Crisp regions have been formally defined on the basis of point sets and point set topology (e.g., [ES97b, Gaa64, Sch97]) which mainly rest on the set operations of union, intersection, and difference. In a straightforward way we will now describe extensions of these two concepts to fuzzy set theory and fuzzy topology.

Fuzzy set theory [Zad65] is an extension and generalization of Boolean set theory. Let X be a classical (crisp) set of objects, called the *universe* (*of discourse*). Membership in a classical subset A of X can then be described by the *characteristic function* $\chi_A : X \to \{0,1\}$ such that for all $x \in X$ holds:

$$\chi_A(x) = \begin{cases} 1 & \text{if and only if } x \in A \\ 0 & \text{if and only if } x \notin A \end{cases}$$

This function, which discriminates sharply between members and non-members of a set, can be generalized such that all elements of X are mapped to the real interval [0,1] indicating the *degree of membership* of these elements in the set in question. Hence, fuzzy set theory permits an element to have partial and multiple membership. Larger values designate higher grades of set membership. Let X again be the universe. Then

$$\mu_{\tilde{A}} : X \to [0,1]$$

is called the *membership function* of $\tilde{A}$, and the set

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \mid x \in X\}$$

is called a *fuzzy set* in X. All elements of X receive a valuation with respect to their membership in $\tilde{A}$. Those elements $x \in X$ that in the classical sense do not belong to $\tilde{A}$ get the membership value $\mu_{\tilde{A}}(x) = 0$; elements $x \in X$ that completely belong to $\tilde{A}$ get the membership value $\mu_{\tilde{A}}(x) = 1$.

There are many ways of extending the set inclusion as well as the basic crisp set operations to fuzzy sets. We will comply with the definitions in [Zad65]. Let $\tilde{A}$ and $\tilde{B}$ be fuzzy sets in X. Then

(i) $\neg\tilde{A} = \{(x, \mu_{\neg\tilde{A}}(x)) \mid x \in X, \mu_{\neg\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x)\}$
(ii) $\tilde{A} \subseteq \tilde{B} \Leftrightarrow \forall\, x \in X : \mu_{\tilde{A}}(x) \le \mu_{\tilde{B}}(x)$
(iii) $\tilde{A} \cap \tilde{B} = \{(x, \mu_{\tilde{A}\cap\tilde{B}}(x)) \mid x \in X \wedge \mu_{\tilde{A}\cap\tilde{B}}(x) = \min(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x))\}$
(iv) $\tilde{A} \cup \tilde{B} = \{(x, \mu_{\tilde{A}\cup\tilde{B}}(x)) \mid x \in X \wedge \mu_{\tilde{A}\cup\tilde{B}}(x) = \max(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x))\}$
(v) $\tilde{A} - \tilde{B} = \tilde{A} \cap \neg\tilde{B}$

A [*strict*] α-*cut* or [*strict*] α-*level set* of a fuzzy set $\tilde{A}$ for a specified value α is the crisp set

$$A_\alpha\,[A^*_\alpha] = \{x \in X \mid \mu_{\tilde{A}}(x) \ge [>]\, \alpha \wedge 0 \le \alpha \le [<]\, 1\}$$

The strict α-cut for $\alpha = 0$ is called *support* of $\tilde{A}$, i.e., $supp(\tilde{A}) = A^*_0$. For a fuzzy set $\tilde{A}$ and $\alpha, \beta \in [0,1]$ holds

(i) $X = A_0$
(ii) $\alpha < \beta \Rightarrow A_\alpha \supseteq A_\beta$

The set of all levels $\alpha \in [0,1]$ that represent distinct α-cuts of a given fuzzy set $\tilde{A}$ is called the *level set* $\Lambda_{\tilde{A}}$ of $\tilde{A}$:

$$\Lambda_{\tilde{A}} = \{\alpha \in [0,1] \mid \exists\, x \in X : \mu_{\tilde{A}}(x) = \alpha\}$$

Fuzzy (point set) topology [Cha68] is a straightforward extension and generalization of ordinary point set topology and allows one to distinguish specific topological structures of a fuzzy set like its closure or interior.

A *fuzzy topology* on a universe X is a family $\tilde{T}$ of fuzzy sets in X satisfying the following conditions:

(i) $X \in \tilde{T}, \emptyset \in \tilde{T}$
(ii) $\tilde{A} \in \tilde{T}, \tilde{B} \in \tilde{T} \Rightarrow \tilde{A} \cap \tilde{B} \in \tilde{T}$
(iii) $\tilde{S} \subseteq \tilde{T} \Rightarrow \bigcup_{\tilde{A} \in \tilde{S}} \tilde{A} \in \tilde{T}$

The pair $(X, \tilde{T})$ is said to be a *fuzzy topological space*. The elements of $\tilde{T}$ are called *open fuzzy sets*. Note that X and $\emptyset$ are crisp sets and simultaneously special fuzzy sets. X corresponds to the fuzzy set $\tilde{X} = \{(x, \mu_X(x)) \mid x \in X \wedge \mu_X(x) = 1\}$. The empty set $\emptyset$ corresponds to the empty fuzzy set $\tilde{\emptyset} = \{(x, \mu_X(x)) \mid x \in X \wedge \mu_X(x) = 0\}$. We will identify X and $\tilde{X}$ as well as $\emptyset$ and $\tilde{\emptyset}$ and use the crisp notations for these two sets.

The family $\tilde{T}'$ of all *closed fuzzy sets* in a fuzzy topological space $(X, \tilde{T})$ is given by

$$\tilde{T}' = \{\neg\tilde{A} \mid \tilde{A} \in \tilde{T}\}$$

The *closure* [*interior*] of a fuzzy set $\tilde{A}$ in a fuzzy topological space $(X, \tilde{T})$ is the smallest closed [largest open] fuzzy set containing $\tilde{A}$ [contained in $\tilde{A}$], i.e.,

$$cl_{\tilde{T}}(\tilde{A}) = \bigcap \{\tilde{S} \mid \tilde{S} \in \tilde{T}' \wedge \tilde{A} \subseteq \tilde{S}\}$$
$$[int_{\tilde{T}}(\tilde{A}) = \bigcup \{\tilde{S} \mid \tilde{S} \in \tilde{T} \wedge \tilde{S} \subseteq \tilde{A}\}]$$

4 Fuzzy Points

Due to our assumption of the point set paradigm, an understanding of the nature of a point, or more precisely a *fuzzy point*, is necessary. There are at least two meaningful definitions for a fuzzy point.

The first definition views a fuzzy point as a point in two-dimensional Euclidean space with a membership value greater than 0, since 0 documents the non-existence of a point. A *fuzzy point* $\tilde{p}$ at (a, b) in $\mathbb{R}^2$, written $\tilde{p}(a, b)$, is a fuzzy singleton in $\mathbb{R}^2$ defined by

$$\mu_{\tilde{p}(a,b)}(x, y) = \begin{cases} m & \text{if } (x, y) = (a, b) \\ 0 & \text{otherwise} \end{cases}$$

with $0 < m \leq 1$. Point $\tilde{p}$ is said to have support (a, b) and value m. Let P_f be the set of all fuzzy points. P_f is, of course, a proper superset of P_c, the set of all crisp points in $\mathbb{R}^2$. For $\tilde{p} = p = (a, b) \in P_c$, we obtain $\mu_{\tilde{p}(a,b)}(x, y) = \chi_p(x, y) = 1$, if $(x, y) = (a, b)$, and 0 otherwise.

The second definition uses a membership function that returns the *degree of proximity* of a point to a reference point $\tilde{p}$. That is, we consider that the point (x, y) is "approximately (a, b)" or "about (a, b)" to the degree $\mu_{\tilde{p}(a,b)}(x, y)$. A fuzzy point $\tilde{p}(a, b)$ is then generally defined by

(i) $\mu_{\tilde{p}(a,b)}$ is upper semicontinuous[1]
(ii) $\mu_{\tilde{p}(a,b)}(x, y) = 1$ if and only if $(x, y) = (a, b)$
(iii) $\forall\ 0 \leq \alpha \leq 1 : \tilde{p}_\alpha$ is a convex[2] subset of $\mathbb{R}^2$

The concrete "distance-based" membership function

$$\mu_{\tilde{p}(a,b)}(x, y) = c^{-\lambda\ ((x-a)^2+(y-b)^2)}$$

with $c \in \mathbb{R}^+$, $c > 1$, and $\lambda > 0$ illustrates this definition. The degree of proximity decreases as (x, y) moves further away from (a, b). It reaches 1 if $(x, y) = (a, b)$.

Unfortunately, this membership function with unbounded support is difficult to represent. Alternatively, we can employ the following, restricted but more practical function which defines a circle around (a, b) with radius $r \in \mathbb{R}^+$:

$$\mu_{\tilde{p}(a,b)}(x, y) = \begin{cases} 1 - \frac{\sqrt{(x-a)^2+(y-b)^2}}{r} & \text{if } (x-a)^2 + (y-b)^2 \leq r^2 \\ 0 & \text{otherwise} \end{cases}$$

Next, we define three *geometric primitives* on fuzzy points which are valid for both definitions of fuzzy points. Let $\tilde{p}(a, b), \tilde{q}(c, d) \in P_f$ with $a, b, c, d \in \mathbb{R}$. Then

(i) $\tilde{p}(a, b) = \tilde{q}(c, d)\ :\Leftrightarrow\ a = c \wedge b = d \wedge \mu_{\tilde{p}(a,b)} = \mu_{\tilde{q}(c,d)}$
(ii) $\tilde{p}(a, b) \neq \tilde{q}(c, d)\ :\Leftrightarrow\ \neg(\tilde{p}(a, b) = \tilde{q}(c, d))$
(iii) $\tilde{p}(a, b)$ and $\tilde{q}(c, d)$ are *disjoint* $:\Leftrightarrow\ supp(\tilde{p}(a, b)) \cap supp(\tilde{q}(c, d)) = \emptyset$

In contrast to crisp points, for fuzzy points we also have a predicate for disjointedness. We are now able to define an object of the fuzzy spatial data type *fpoint* as a set of disjoint fuzzy points:

$$fpoint = \{Q \subseteq P_f \mid \forall\ \tilde{p}, \tilde{q} \in Q : \tilde{p}(a, b) \text{ and } \tilde{q}(c, d) \text{ are } disjoint \wedge Q \text{ is finite}\}$$

5 Fuzzy Lines

In this section we define a data type for *fuzzy lines*. For that, we first introduce a *simple* fuzzy line as a continuous curve with smooth transitions of membership grades between neighboring points of the line. We assume a total order on $\mathbb{R}^2$ which is given by the lexicographic order "$<$" on the coordinates (first x, then y). The membership function of a *simple fuzzy line* $\tilde{l}$ is then defined by

$$\mu_{\tilde{l}} : f_{\tilde{l}} \rightarrow [0, 1] \text{ with } f_{\tilde{l}} : [0, 1] \rightarrow \mathbb{R}^2 \text{ such that}$$

[1] A function $f : X \rightarrow \mathbb{R}$ is *upper semicontinuous* $:\Leftrightarrow \forall\ r \in \mathbb{R} : \{x \mid f(x) < r\}$ is open.
[2] A set $X \subseteq \mathbb{R}^2$ is called *convex* $:\Leftrightarrow \forall\ p, q \in \mathbb{R}^2\ \forall\ \lambda \in \mathbb{R}^+$ with $0 < \lambda < 1 : r = \lambda\, p + (1 - \lambda)\, q \in X$ (p, q, and r are here regarded as vectors)

(i) $\mu_{\tilde{l}}$ is continuous
(ii) $f_{\tilde{l}}$ is continuous
(iii) $\forall\ a, b \in (0,1) : a \neq b \ \Rightarrow\ f_{\tilde{l}}(a) \neq f_{\tilde{l}}(b)$
(iv) $\forall\ a \in \{0,1\}\ \forall\ b \in (0,1) : f_{\tilde{l}}(a) \neq f_{\tilde{l}}(b)$
(v) $f_{\tilde{l}}(0) < f_{\tilde{l}}(1) \vee (f_{\tilde{l}}(0) = f_{\tilde{l}}(1) \wedge \forall a \in (0,1) : f_{\tilde{l}}(0) < f_{\tilde{l}}(a))$

Function $f_{\tilde{l}}$ on its own models a continuous, simple *crisp* line (a *simple curve*). The points $f_{\tilde{l}}(0)$ and $f_{\tilde{l}}(1)$ are called the *end points* of f. The definition allows loops ($f_{\tilde{l}}(0) = f_{\tilde{l}}(1)$) but prohibits equality of interior points and thus self-intersections (condition (iii)) and equality of an interior with an end point (condition (iv)). The last condition ensures uniqueness of representation, i.e., in a *closed simple line* $f_{\tilde{l}}(0)$ must be the leftmost point.

Let S be the set of fuzzy simple lines, and let $T \subseteq S$. An *S-complex C over T* is a finite set $C = \{\tilde{l}_1, \ldots, \tilde{l}_n\} \subseteq T$ such that[3]

(i) $\forall\ 1 \leq i < j \leq n : f_{\tilde{l}_i}((0,1)) \cap f_{\tilde{l}_j}((0,1)) = \emptyset$
(ii) $\forall\ 1 \leq i < j \leq n : \{f_{\tilde{l}_i}(0), f_{\tilde{l}_i}(1)\} \cap f_{\tilde{l}_j}((0,1)) = \emptyset$
(iii) $\forall\ 1 \leq i \leq n\ \exists\ 1 \leq j \leq n, j \neq i : \{f_{\tilde{l}_i}(0), f_{\tilde{l}_i}(1)\} \cap \{f_{\tilde{l}_j}(0), f_{\tilde{l}_j}(1)\} \neq \emptyset$
(iv) For all $1 \leq i,j \leq n$ and for all $a, k \in \{0,1\}$ let $V^a_{\tilde{l}_i} = \{(j,k) \mid f_{\tilde{l}_i}(a) = f_{\tilde{l}_j}(k)\}$. Then we require: $\forall\ 1 \leq i \leq n\ \forall\ a \in \{0,1\} : (|V^a_{\tilde{l}_i}| = 1) \vee (|V^a_{\tilde{l}_i}| > 2)$
(v) $\forall\ 1 \leq i \leq n\ \forall\ a \in \{0,1\}\ \forall\ (j,k) \in V^a_{\tilde{l}_i} : \mu_{\tilde{l}_i}(f_{\tilde{l}_i}(a)) = \mu_{\tilde{l}_j}(f_{\tilde{l}_j}(k))$

Condition (i) requires that the elements of an S-complex do not intersect or overlap within their interior. Moreover, they may not be touched within their interior by an endpoint of another element (condition (ii)). Condition (iii) ensures the property of connectivity of an S-complex; isolated fuzzy simple lines are disallowed. Condition (iv) expresses that each endpoint of an element of C must belong to exactly one or more than two incident elements of C (note that always $(i,a) \in V^a_{\tilde{l}_i}$). This condition supports the requirement of maximal elements and hence achieves minimality of representation. Condition (v) requires that the membership values of more than two elements of C with a common end point must have the same membership value; otherwise we get a contradiction saying that a point of an S-complex has more than one different membership value.

All conditions together define an S-complex over T as a *connected planar fuzzy graph* with a unique representation. The corresponding point set of C is $points(C) = \bigcup_{\tilde{l} \in C} f_{\tilde{l}}([0,1])$. The set of all S-complexes over T is denoted by $SC(T)$. The disjointedness of any two S-complexes $C_1, C_2 \in SC(T)$ is defined as follows:

$$C_1 \text{ and } C_2 \text{ are } disjoint :\Leftrightarrow points(C_1) \cap points(C_2) = \emptyset$$

A fuzzy spatial data type for fuzzy lines called *fline* can now be defined in two equivalent ways. The "structured view" is based on S-complexes:

$$fline = \{D \subseteq SC(S) \mid \forall\ C_1, C_2 \in D : C_1 \text{ and } C_2 \text{ are } disjoint \wedge D \text{ is finite}\}$$

[3] The application of a function f to a set X of values is defined as $f(X) = \{f(x) \mid x \in X\}$.

Let $1_r = \mathbb{R}^2 \times \{1\}$. The "flat view" emphasizing the point set paradigm is:

$$\mathit{fline} = \{\tilde{Q} \subseteq 1_r \mid \exists\, D \subseteq SC(S) : \bigcup_{C \in D} \mathit{points}(C) = \mathit{supp}(\tilde{Q})\}$$

6 Fuzzy Regions

The aim of this section is to develop and formalize the concept of a *fuzzy region*. Section 6.1 informally discusses the intrinsic features of fuzzy regions, classifies them, gives application examples for them, and compares them to classical crisp regions. After this motivation, Section 6.2 provides their formal definition. Finally, Section 6.3 gives examples of possible membership functions for them.

6.1 What are Fuzzy Regions?

The question what a *crisp region* is has been treated in many publications. A very general definition defines a crisp region as a set of disjoint, connected areal components, called *faces*, possibly with disjoint *holes* [ES97b, GS95, Sch97] in the Euclidean space $\mathbb{R}^2$. This model has the nice property that it is closed under (appropriately defined) geometric union, intersection, and difference operations. It allows crisp regions to contain holes and islands within holes to any finite level.

By analogy with the generalization of crisp sets to fuzzy sets, we strive for a generalization of crisp regions to fuzzy regions on the basis of the point set paradigm and fuzzy concepts. At the same time we would like to transfer the structural definition of crisp regions (i.e., the component view) to fuzzy regions. Thus, the structure of a fuzzy region is supposed to be the same as for a crisp region but with the exception and generalization which amounts to a relaxation and hence greater flexibility of the strict belonging or non-belonging principle of a point in space to a specific region and which enables a partial membership of a point in a region. This is just what the term "fuzzy" means here.

There are at least three possible, related interpretations for a point in a fuzzy region. First, this situation may be interpreted as the *degree of belonging* to which that point is *inside* or *part of* some areal feature. Consider the transition between a mountain and a valley and the problem to decide which points have to be assigned to the valley and which points to the mountain. Obviously, there is no strict boundary between them, and it seems to be more appropriate to model the transition by partial and multiple membership. Second, this situation may indicate the *degree of compatibility* of the individual point with the attribute or concept represented by the fuzzy region. An example are "warm areas" where we must decide for each point whether and to which grade it corresponds to the concept "warm". Third, this situation may be viewed as the *degree of concentration* of some attribute associated with the fuzzy region at the particular point. An example is air pollution where we can assume the highest concentration at power stations, for instance, and lower concentrations with increasing distance from them. All these related interpretations give evidence of *fuzziness*.

When dealing with crisp regions, the user usually does not employ point sets as a method to conceptualize space. The user rather thinks in terms of sharply determined *boundaries* enclosing and grouping areas with *equal* properties or attributes and separating different regions with different properties from each other; he or she has purely *qualitative* concepts in mind. This view changes when fuzzy regions come into play. Besides the qualitative aspect, in particular the *quantitative* aspect becomes important, and boundaries in most cases disappear (between a valley and a mountain there is no strict boundary!). The distribution of attribute values within a region and transitions between different regions may be *smooth* or *continuous*. This feature just characterizes fuzzy regions.

We now give a classification of fuzzy regions from an application point of view. The classification extends from fuzzy regions with highest vagueness and lowest gradation of attribute values to fuzzy regions with lowest vagueness and highest gradation of attribute values. The given application examples are basically valid for each class. How to model areal features as fuzzy regions depends on the application and on the "preciseness" and quality of information.

Core-Boundary Fuzzy Regions. If there is only insufficient knowledge about the grade of indeterminacy of the vague parts of a region, a first approach is to differentiate between its *core*, its *boundary*, and its *exterior* which relate to those parts that definitely belong, *perhaps* belong, and definitely do not belong, respectively, to the region. This extension just corresponds to the approach of *vague regions* where core and boundary are modeled by crisp regions. It can also be simply modeled by a fuzzy region by assigning the membership function value 1 to each point of the core, value 0 to each point of the exterior, and value $\frac{1}{2}$ (halfway between completely true and completely false) to each point of the boundary. It is important to note that a boundary in this sense can be a region and has thus a different and generalized meaning compared to traditional, crisp boundaries[4]. We will denote fuzzy regions based on a three-valued logic as *core-boundary (fuzzy) regions*.

An application example is a lake which has a minimal water level in dry periods (core) and a maximal water level in rainy periods (boundary given as the difference between maximal and minimal water level). Dry periods can entail puddles. Small islands in the lake which are less flooded by water in dry and more (but never completely) flooded in rainy periods can be modeled through holes surrounded by a boundary. If an island like a sandbank can be flooded completely, it belongs to the boundary part.

Finite-Valued Fuzzy Regions. The next step lifts the restriction of having only one degree of fuzziness. The introduction of different degrees leads from fuzzy regions based on a three-valued logic to fuzzy regions based on a *finite-valued* and thus *multivalued logic*. This enables us to describe more precisely the

[4] Nevertheless, core, boundary, and exterior are separated from each other by ordinary, strict "boundaries" as we know them from ordinary point set topology.

degree of membership of a point in a fuzzy region. The membership function value $\frac{3}{4}$ ($\frac{1}{4}$) could express that it is mostly true (false) and only a little false (true) that a point is an element of a specific fuzzy region. We will call this kind of fuzzy regions *finite-valued (fuzzy) regions*. If $n \in \mathbb{N}$ is the number of possible "truth values", an n-valued membership function turns out to be quite useful for representing a wide range of belonging of a point to a fuzzy region.

An application example are regions of different possibilities for virus infections. Regions could be categorized by n different risk levels extending from areas with extreme risk of infection over areas with average risk of infection to safe areas.

The two classes of fuzzy regions described so far have predominantly a *qualitative* character. This means, that the numbers involved in membership functions of a fuzzy region only play a symbolic role and that their size is of lower importance. Essentially, a total and bijective mapping is defined between n possible categories expressing different degrees of fuzziness and n discrete values out of the range $[0, 1]$. Although the selection of the n discrete values is arbitrary (they only must be disjoint from each other, and there is no order needed between them), they are usually chosen in a way that agrees with our intuition.

Interval-Based Fuzzy Regions. The following two classes emphasize a more *quantitative* character of fuzzy regions. Consider an ordered set of n arbitrary but disjoint values of the interval $[0, 1]$ and the assignment of exactly one of these values, let us say, v, to all points of a specific connected component c (a face) of a fuzzy region. We can then interpret such a value v for all points of c as their guaranteed minimal degree of belonging to c. Hence, v represents a lower bound. Since the set of values is ordered, each value v (except for the highest value) has a successor w with respect to the defined order, i.e., $v < w$. This implies that no point of c can have a value greater than w, since otherwise these points would have to be labeled with the value w. This justifies to implicitly map all points of c to the label $[v, w]$, i.e., to a closed interval. The meaning is that the degree of membership of each point of c is somewhere between v and w (we do not have more information). We denote this kind of fuzzy regions as *interval-based (fuzzy) regions*. Each pair of the $n - 1$ possible intervals is either disjoint or adjacent with common bounds. All intervals together form a finite covering of the unit interval $[0, 1]$.

An application example is a map about the population density of a country. According to a predefined interval classification, the country is subdivided into regions showing the minimal guaranteed population density per km^2 for each region. The density values of different regions can be rather different. Another example are weather maps on television which usually show single reference temperatures as sample data spread over the map and representing temperature zones. Here we assume that a direct path from a lower to a higher reference temperature is accompanied by smoothly increasing temperatures. Transitions between different regions are here smooth.

Smooth Fuzzy Regions. A last and very important class of fuzzy regions, which has so far not been treated in the literature, takes advantage of available knowledge about the distribution of attribute values within a fuzzy region. This knowledge can be gained by an expert through appropriate membership functions. We require that the distribution of attribute values within a fuzzy region is *smooth* (with a finite number of exceptions). This can be achieved by so-called *predominantly continuous* membership functions. We call this kind of fuzzy regions *predominantly smooth (fuzzy) regions*. As a special case we obtain *(totally) smooth (fuzzy) regions* with no continuity gaps.

There are a lot of spatial phenomena showing a smooth behavior. Application examples are air pollution (Figure 1), temperature zones, magnetic fields, storm intensity, and sun insolation. Predominantly smooth regions are the most general class of fuzzy regions and comprise all other aforementioned classes which are obviously (predominantly) continuous. This especially means that combinations of different classes are possible without any problems.

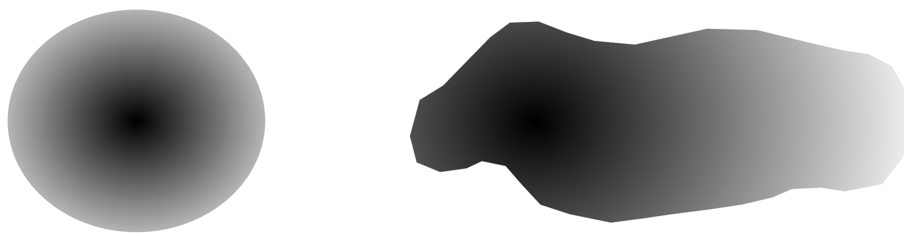

Fig. 1. This figure demonstrates a possible visualization of a fuzzy region which could model the expansion of air pollution caused by a power station. The left image shows a radial expansion where the degree of pollution concentrates in the center (darker locations) and decreases with increasing distance from the power station (brighter locations). The right image has the same theme but this time we imagine that the power station is surrounded by high mountains to the north, the south, and the west. Hence, the pollution cannot escape in these directions and finds its way out of the valley in eastern direction. In both cases we can recognize the smooth transitions to the exterior.

6.2 Formal Definition of Fuzzy Regions

Since our objective is to model two-dimensional fuzzy areal objects for spatial applications, we consider a fuzzy topology $\tilde{T}$ on the Euclidean space (plane) $\mathbb{R}^2$. In this spatial context we denote the elements of $\tilde{T}$ as *fuzzy point sets*. The membership function for a fuzzy point set $\tilde{A}$ in the plane is then described by $\mu_{\tilde{A}} : \mathbb{R}^2 \to [0, 1]$.

From an application point of view, there are two observations that prevent a definition of a fuzzy region simply as a fuzzy point set. We will discuss them now in more detail and at the same time elaborate properties of fuzzy regions.

Avoiding Geometric Anomalies: Regularization. The first observation refers to a necessary *regularization* of fuzzy point sets. The first reason for this measure is that fuzzy (as well as crisp) regions that actually appear in spatial applications in most cases cannot be just modeled as arbitrary point sets but have to be represented as point sets that do not have "geometric anomalies" and that are in a certain sense *regular*. Geometric anomalies relate to isolated or dangling line or point features and missing lines and points in the form of cuts and punctures. Spatial phenomena with such degeneracies never appear as entities in reality. The second reason is that, from a data type point of view, we are interested in fuzzy spatial data types that satisfy closure properties for (appropriately defined) geometric union, intersection, and difference.

We are, of course, confronted with the same problem in the crisp case where the problem can be avoided by the concept of *regularity* [ES97b, Sch97, Til80]. It turns out to be useful to appropriately transfer this concept to the fuzzy case. Let $\tilde{A}$ be a fuzzy set of a fuzzy topological space $(\mathbb{R}^2, \tilde{T})$. Then

$\tilde{A}$ is called a *regular open fuzzy set* if $\tilde{A} = int_{\tilde{T}}(cl_{\tilde{T}}(\tilde{A}))$

Whereas crisp regions are usually modeled as *regular closed crisp sets*, we will use *regular open fuzzy sets* due to their vagueness and their usual lack of boundaries. Regular open fuzzy sets avoid the aforementioned geometric anomalies, too. Since application examples show that fuzzy regions can also be partially bounded, we admit *partial boundaries* with a crisp or fuzzy character. For that purpose we define the following fuzzy set:

$$frontier_{\tilde{T}}(\tilde{A}) := \{((x, y), \mu_{\tilde{A}}(x, y)) \mid (x, y) \in supp(\tilde{A}) - supp(int_{\tilde{T}}(\tilde{A}))\}$$

A fuzzy set $\tilde{A}$ is now called a *spatially regular fuzzy set* iff

(i) $int_{\tilde{T}}(\tilde{A})$ is a regular open fuzzy set
(ii) $frontier_{\tilde{T}}(\tilde{A}) \subseteq frontier_{\tilde{T}}(cl_{\tilde{T}}(int_{\tilde{T}}(\tilde{A})))$
(iii) $frontier_{\tilde{T}}(\tilde{A})$ is a partition of n connected boundary parts (fuzzy sets)

We can conclude that $frontier_{\tilde{T}}(\tilde{A}) = \emptyset$ if $\tilde{A}$ is regular open. We will base our definition of fuzzy regions on spatially regular fuzzy sets and define a *regularization* function reg_f which associates the interior of a fuzzy set $\tilde{A}$ with its corresponding regular open fuzzy set and which restricts the partial boundary of $\tilde{A}$ (if it exists at all) to a part of the boundary of the corresponding regular closed fuzzy set of $\tilde{A}$:

$$reg_f(\tilde{A}) := int_{\tilde{T}}(cl_{\tilde{T}}(\tilde{A})) \cup (frontier_{\tilde{T}}(\tilde{A}) \cap frontier_{\tilde{T}}(cl_{\tilde{T}}(int_{\tilde{T}}(\tilde{A}))))$$

The different components of the regularization process work as follows: the *interior* operator $int_{\tilde{T}}$ eliminates dangling point and line features since their interior is empty. The *closure* operator $cl_{\tilde{T}}$ removes cuts and punctures by appropriately adding points. Furthermore, the *closure* operator introduces a fuzzy boundary (similar to a crisp boundary in the ordinary point-set topological sense) separating the points of a closed set from its exterior. The operator $frontier_{\tilde{T}}$ supports the restriction of the boundary.

The following statements about set operations on regular open fuzzy sets are given informally and without proof. The intersection of two regular open fuzzy sets is regular open. The union, difference, and complement of two regular open fuzzy sets are not necessarily regular open since they can produce anomalies. Correspondingly, this also holds for spatially regular fuzzy sets. Hence, we introduce *regularized set operations* on spatially regular fuzzy sets that preserve regularity. Let $\tilde{A}$, $\tilde{B}$ be spatially regular fuzzy sets of a fuzzy topological space $(\mathbb{R}^2, \tilde{T})$, and let $a \dot{-} b = a - b$ for $a \geq b$ and $a \dot{-} b = 0$ otherwise $(a, b \in \mathbb{R}_0^+)$. Then

(i) $\tilde{A} \cup_r \tilde{B} := reg_f(\tilde{A} \cup \tilde{B})$

(ii) $\tilde{A} \cap_r \tilde{B} := reg_f(\tilde{A} \cap \tilde{B})$

(iii) $\tilde{A} -_r \tilde{B} := reg_f(\{((x, y), \mu_{\tilde{A} -_r \tilde{B}}(x, y) \mid (x, y) \in \tilde{A} \ \wedge$
$\mu_{\tilde{A} -_r \tilde{B}}(x, y) = \mu_{\tilde{A}}(x, y) \dot{-} \mu_{\tilde{B}}(x, y)\})$

(iv) $\neg_r \tilde{A} := reg_f(\neg \tilde{A})$

Note that we have changed the meaning of difference (i.e., $\tilde{A} -_r \tilde{B} \neq \tilde{A} \cap_r \neg \tilde{B}$) since the right side of the inequality does not seem to make great sense in the spatial context. Regular open fuzzy sets, spatially regular fuzzy sets, and regularized set operations express a natural formalization of the desired dimension-preserving property of set operations. In the crisp case this is taken for granted but mostly never fulfilled by spatial type systems, geometric algorithms, spatial database systems, and GIS.

Whereas the subspace *RCCS* of *r*egular *c*losed *c*risp *s*ets together with the crisp regular set operations "$\cup$" and "$\cap$" and the set-theoretic order relation "$\subseteq$" forms a Boolean lattice [ES97b], this is not the case for *SRFS* denoting the subspace of *s*patially *r*egular *f*uzzy *s*ets. Here we obtain the (unproven but obvious) statement that *SRFS* together with the regularized set operations "$\cup_r$" and "$\cap_r$" and the fuzzy set-theoretic order relation "$\subseteq$" is a pseudo-complemented distributive lattice.

This implies that (i) (*SRFS*, $\subseteq$) is a partially ordered set (reflexivity, antisymmetry, transitivity), (ii) every pair $\tilde{A}$, $\tilde{B}$ of elements of *SRFS* has a least upper bound $\tilde{A} \cup_r \tilde{B}$ and a greatest lower bound $\tilde{A} \cap_r \tilde{B}$, (iii) (*SRFS*, $\subseteq$) has a maximal element $1_r := \{((x, y), \mu(x, y)) \mid (x, y) \in \mathbb{R}^2 \wedge \mu(x, y) = 1\}$ (identity of "$\cap_r$") and a minimal element $0_r := \{((x, y), \mu(x, y)) \mid (x, y) \in \mathbb{R}^2 \wedge \mu(x, y) = 0\}$ (identity of "$\cup_r$"), and (iv) algebraic laws like idempotence, commutativity, associativity, absorption, and distributivity hold for "$\cup_r$" and "$\cap_r$".

(*SRFS*, $\subseteq$) is not a complementary lattice. Although the algebraic laws of involution and dualization hold, this is not true for the laws of complementarity. If we take the standard fuzzy set operations presented in Section 3 as a basis, the law of excluded middle $\tilde{A} \cup_r \neg \tilde{A} = 1_r$ and the law of contradiction $\tilde{A} \cap_r \neg \tilde{A} = 0_r$ do *not* hold in general. This fact explains the term "pseudo-complemented" from above and is no weakness of the model but only an indication of fuzziness.

Modeling Smooth Attribute Changes: Predominantly Continuous Membership Functions. The second observation is that according to the application cases shown in Section 6.1 the mapping $\mu_{\tilde{A}}$ itself may not be arbitrary but must take into account the intrinsic smoothness of fuzzy regions. This property can be modeled by the well known mathematical concept of *continuity* and results in special continuous membership functions for fuzzy regions. We say that a function f contains a *continuity gap* at a point x_0 of its domain if f is semicontinuous but not continuous at x_0. Function f is called *predominantly continuous* if f is continuous and has at most a *finite* number of continuity gaps.

Defining Fuzzy Regions. The type *fregion* for fuzzy regions can now be defined in the following way:

$$\mathit{fregion} = \{\tilde{R} \in \mathit{SRFS} \mid \mu_{\tilde{R}} \text{ is predominantly continuous}\}$$

6.3 Examples of Membership Functions for Fuzzy Regions

In this section we give some simple examples of membership functions which fulfil the properties required in Section 6.2. The determination of suitable membership functions is the difficulty in using the fuzzy set approach. Frequently, expert and empirical knowledge is necessary and used to design appropriate functions. We start with an example for a smooth fuzzy region. By taking a crisp region A with boundary B_A as a reference object, we can construct a fuzzy region on the basis of the following distance-based membership function:

$$\mu_{\tilde{A}} = \begin{cases} 1 & \text{if } (x,y) \in A \\ a^{-\lambda\, d((x,y),B_A)} & \text{if } (x,y) \notin A \end{cases}$$

where $a \in \mathbb{R}^+$ and $a > 1$, $\lambda \in \mathbb{R}^+$ is a constant, and $d((x,y), B_A)$ computes the distance between point (x,y) and boundary B_A in the following way:

$$d((x,y), B_A) = \min\{\mathit{dist}((x,y),(x',y')) \mid (x',y') \in B_A\}$$

where $\mathit{dist}(p,q)$ is the usual Euclidean distance between two points $p, q \in \mathbb{R}^2$. Unfortunately, this membership function leads to an unbounded spatially regular fuzzy set (regular open fuzzy set) which is impractical for implementation. We can also give a similar definition of a membership function with bounded support:

$$\mu_{\tilde{A}} = \begin{cases} 1 & \text{if } (x,y) \in A \\ a^{1-\frac{1}{\lambda} d((x,y),B_A)} & \text{if } (x,y) \notin A,\ d((x,y),B_A) \le \lambda \\ 0 & \text{otherwise} \end{cases}$$

In the same way as the distance from a point outside of A to B_A increases to λ, the degree of membership of this point to $\tilde{A}$ decreases to zero.

[Use96] also presents membership functions for smooth fuzzy regions. The applications considered are air pollution defined as a fuzzy region with membership values based on the distance from a city center and a hill with elevation as the controlling value for the membership function. [LAB96] models the transition of two smooth regions for soil units with symmetric membership functions.

A method to design a membership function for a finite-valued region with n possible membership values (truth values) is to code the n values by rational numbers in the unit interval $[0, 1]$. For that purpose, the unit interval is evenly divided into $n-1$ subintervals and takes their endpoints as membership values. We obtain the set $T_n = \{\frac{i}{n-1} \mid n \in \mathbb{N}, 0 \leq i \leq n-1\}$ of truth values. Assuming that we intend to model air pollution caused by a power station located at point $p \in \mathbb{R}^2$, we can define the following (simplified) membership function for $n = 5$ degrees of truth representing, for instance, areas of extreme, high, average, low, and no pollution ($a, b, c, d \in \mathbb{R}^+$):

$$\mu_{\tilde{A}}(x, y) = \begin{cases} 1 & \text{if } dist(p, (x, y)) \leq a \\ \frac{3}{4} & \text{if } a < dist(p, (x, y)) \leq b \\ \frac{1}{2} & \text{if } b < dist(p, (x, y)) \leq c \\ \frac{1}{4} & \text{if } c < dist(p, (x, y)) \leq d \\ 0 & \text{if } d < dist(p, (x, y)) \end{cases}$$

7 Structured Views of Fuzzy Regions

The formal definition of a fuzzy region given in Section 6.2 is conceptually somehow "structureless" in the sense that only "flat" point sets are considered and no structural information is revealed. In the following four subsections some "semantically richer" characterizations of fuzzy regions are presented which enable a better understanding of fuzzy regions. On the one hand they subdivide fuzzy regions into fuzzy components and on the other hand they describe them as collections of crisp regions. Moreover, they give hints for a possible implementation.

7.1 Fuzzy Regions as Multi-Component Objects

The first structured view considers a fuzzy region as a set of *fuzzy components*. For a definition we need a notion of *connectedness* for fuzzy regions. A *separation* of a fuzzy region $\tilde{Y}$ is a pair $\tilde{A}, \tilde{B}$ of fuzzy subregions satisfying the following four conditions:

(i) $\tilde{A} \neq \emptyset, \tilde{B} \neq \emptyset$
(ii) $\tilde{Y} = \tilde{A} \cup_r \tilde{B}$
(iii) $\tilde{A} \cap int_{\tilde{T}}(\tilde{B}) = \emptyset \wedge int_{\tilde{T}}(\tilde{A}) \cap \tilde{B} = \emptyset$
(iv) $|\tilde{A} \cap_r \tilde{B}|$ is finite

If a separation of $\tilde{Y}$ into $\tilde{A}$ and $\tilde{B}$ exists, then $\tilde{Y}$ is said to be *separated*, and we call $\tilde{A}$ and $\tilde{B}$ to be *disjoint*. Otherwise $\tilde{Y}$ is said to be *connected*. Note that condition (iii) of the definition uses the usual fuzzy intersection operation and not the one on spatially regular fuzzy sets since the latter requires two fuzzy regions as operands. The property of disjointedness (condition (iv)) requires that the two fuzzy subregions $\tilde{A}$ and $\tilde{B}$ may at most share a finite number of boundary points; this makes sense since otherwise they could be simply merged into one fuzzy subregion. We now continue this separation process and decompose a fuzzy region $\tilde{Y}$ into its maximal set of pairwise disjoint fuzzy components $\tilde{Y} = \{\tilde{A}_1, \ldots, \tilde{A}_n\}$ (in the spatial context this decomposition is always finite) so that we obtain with $I = \{1, \ldots, n\}$:

(i) $\forall\, i \in I : \tilde{A}_i \neq \tilde{\emptyset}$
(ii) $\tilde{Y} = \bigcup_{r\, i \in I} \tilde{A}_i$
(iii) $\forall\, i, j \in I, i \neq j : \tilde{A}_i \cap int_{\tilde{T}}(\tilde{A}_j) = \emptyset \wedge int_{\tilde{T}}(\tilde{A}_i) \cap \tilde{A}_j = \emptyset$
(iv) $\forall\, i, j \in I, i \neq j : |\tilde{A}_i \cap_r \tilde{A}_j|$ is finite
(v) $\forall\, i \in I : (\tilde{A}_i$ is connected $\wedge\ \nexists\ \tilde{B} \supset \tilde{A}_i : \tilde{B}$ is connected$)$

We call each fuzzy component $\tilde{A}_i$ a *fuzzy face*. Hence, we obtain:

A *fuzzy region* is a set of pairwise disjoint *fuzzy faces*.

A question arises whether also *fuzzy holes* can be identified from the point set view of a fuzzy region. This question has to be negated. Let us briefly consider the crisp case. If A is a crisp region, its faces can have holes which belong to the complement (exterior) of A, i.e., to $\mathbb{R}^2 - A$, and are "enclosed" by A. Unfortunately, ordinary point set topology offers no method to extract holes from a (regular closed) point set as separate components; they are simply part of the complement. Note that this does not mean that regions with holes cannot be modeled. Some research work in [ECF94, Sch97, WB93], for example, shows that this is possible by selecting a constructive approach. Roughly speaking, the idea is to assume that the holes of A are already given as regions and to subtract these holes from a "generalized region A*" being isomorphic to a closed disc and being the union of A and the holes. But since this a pure set operation, afterwards A "forgets" how it was produced and cannot reconstruct its past. Similarly to the crisp case, holes cannot be identified from a (spatially regular) fuzzy point set, since fuzzy topology also offers no concept of holes.

Moreover, we are here faced with the problem of the nature of a fuzzy hole. By analogy with the crisp case, we could say that the fuzzy holes of a fuzzy region $\tilde{A}$ exclusively contain all points that are enclosed by any fuzzy face of $\tilde{A}$ and that have membership grade 0 in $\tilde{A}$. But then, a fuzzy hole is crisp and a subset of the set

$$H = \{((x, y), 1) \mid (x, y) \in supp(\neg\tilde{A})\}$$

This model of a fuzzy hole is unsatisfactory in the sense that it only deals with those points enclosed by $\tilde{A}$ that definitely do not belong to $\tilde{A}$. It does not take into account the complement of those points of $\tilde{A}$ belonging only partially to $\tilde{A}$, i.e., the model does not consider the set

$$\underline{\tilde{A}} = \{((x, y), m) \mid (x, y) \in supp(\tilde{A}) \wedge m = 1 - \mu_{\tilde{A}}(x, y)\}$$

called the *anti-fuzzy region* of $\tilde{A}$.

One could argue that the points of $\underline{\tilde{A}}$ also belong to the fuzzy holes. And indeed, we will take this view. The consequence is that for a fuzzy face there exists exactly one fuzzy hole.

7.2 Fuzzy Regions as Three-Part Crisp Regions

The second structured view leads to a simplification of an originally smooth fuzzy region to a core-boundary region and thus to a change from a quantitative to a qualitative perspective. It distinguishes between the kernel, the boundary, and the exterior as the three parts of a fuzzy region. For a fuzzy region $\tilde{A}$, these parts are defined as crisp regions (regular closed sets)[5]:

$$\begin{aligned} kernel(\tilde{A}) &= reg_c(\{(x,y) \in \mathbb{R}^2 \mid \mu_{\tilde{A}}(x,y) = 1\}) \\ boundary(\tilde{A}) &= reg_c(\{(x,y) \in \mathbb{R}^2 \mid 0 < \mu_{\tilde{A}}(x,y) < 1\}) \\ exterior(\tilde{A}) &= reg_c(\{(x,y) \in \mathbb{R}^2 \mid \mu_{\tilde{A}}(x,y) = 0\}) \end{aligned}$$

The *kernel* identifies the part that definitely belongs to $\tilde{A}$. The *exterior* determines the part that definitely does not belong to $\tilde{A}$. The indeterminate character of $\tilde{A}$ is summarized in the *boundary* of $\tilde{A}$ in a unified and simplified manner. Kernel and boundary can be adjacent with a common border, and kernel and/or boundary can be empty. This view corresponds exactly to the already described concept of *vague regions* with its three-valued logic [ES97b].

All in all, this view presents only a very coarse and restricted description of fuzzy regions since it differentiates only between three parts. The original gradation in the membership values of the points of the boundary gets lost. The benefit of this view lies in the implementation since efficient representation methods and algorithms for crisp regions can be used.

7.3 Fuzzy Regions as Collections of Crisp α-Level Regions

The third structured view attempts to diminish the drawbacks of the three-part view of fuzzy regions and to avoid the great information loss in this representation. It describes a fuzzy region in terms of nested α-level sets. Let $\tilde{A}$ be a fuzzy region. Then we represent a region A_a for an $\alpha \in [0,1]$ as

$$A_a = reg_c(\{(x,y) \in \mathbb{R}^2 \mid \mu_{\tilde{A}}(x,y) \geq \alpha\})$$

We call A_a an *α-level region*. Clearly, A_a is a crisp region whose boundary is defined by all points with membership value α. Note that A_a can have holes. The kernel of $\tilde{A}$, as it has been defined in Section 7.2, is then equal to $A_{1.0}$. A property of the α-level regions of a fuzzy region is that they are nested, i.e., if

[5] Correspondingly, for a crisp set A, the regularization function reg_c is defined as $reg_c(A) = cl_T(int_T(A))$ where T is a topology for a universe X and cl_T and int_T are the *closure* and *interior* operators on a topological space (X,T).

we select membership values $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_n > \alpha_{n+1} = 0$ for some $n \in \mathbb{N}$, then

$$A_{\alpha_1} \subseteq A_{\alpha_2} \subseteq \cdots \subseteq A_{\alpha_n} \subseteq A_{\alpha_{n+1}}$$

We here describe the finite, discrete case that enables us to model and implement finite-valued and interval-based regions. If $\Lambda_{\tilde{A}}$ is infinite, then there are obviously infinitely many α-level regions which can only be finitely represented within this view if we make a finite selection of values. In the discrete case, if $|\Lambda_{\tilde{A}}| = n + 1$ and we take all these occurring membership values of a fuzzy region, we can replace "$\subseteq$" by "$\subset$" in the inclusion relationships above. This follows from the fact that for any $p \in A_{\alpha_i} - A_{\alpha_{i-1}}$ with $i \in \{2, \ldots, n+1\}$, $\mu_{\tilde{A}}(p) = \alpha_i$. For the continuous case, we get $\mu_{\tilde{A}}(p) \in [\alpha_i, \alpha_{i-1})$ which leads to interval-based regions. As a result, we obtain:

A *fuzzy region* is a (possibly infinite) set of α-level regions, i.e., $\tilde{A} = \{A_{\alpha_i} \mid 1 \le i \le |\Lambda_{\tilde{A}}|\}$ with $\alpha_i > \alpha_{i+1} \Rightarrow A_{\alpha_i} \subseteq A_{\alpha_{i+1}}$ for $1 \le i \le |\Lambda_{\tilde{A}}| - 1$.

From the implementation perspective, one of the advantages of using (a finite collection of) α-level sets to describe fuzzy regions is that existing geometric data structures and geometric algorithms known from Computational Geometry [PS85] can be applied.

7.4 Fuzzy Regions as α-Partitions

The fourth structured view is partially motivated by the previous one and describes a fuzzy region as a partition. A partition in the spatial context, called a *spatial partition* [ES97a], is a subdivision of the plane into pairwise disjoint (crisp) regions (called *blocks*) where each block is associated with an attribute and where adjacent blocks are not allowed to be labeled with the same attribute. It differs from the set-theoretic notion of a partition in the sense that it, of course, relates to space and that it incorporates a treatment of common boundary points which at the same time may belong to two adjacent blocks.

From an application point of view, different blocks of a spatial partition are often marked differently, i.e., different *labels* of some set L are assigned to different blocks. Thus, in a certain way, L determines the type of a partition. This leads to spatial partitions of type L that are functions $\pi : \mathbb{R}^2 \to L$. In most cases, partitions are defined only partially, i.e., there are blocks (frequently called the *exterior* of a partition) which have no explicitly assigned labels. To complete π to a total function, we assume a label $\perp_L$ (called *undefined* or *unknown*) for each label type L and require that the exterior of a partition is labeled by $\perp_L$.

Like for crisp regions, we also desire regularity for the blocks of a spatial partition. We require the interiors of blocks to be regular open sets. Since points on the boundary cannot be uniquely assigned to either adjacent block, we cannot simply map them to single L-values. Instead, boundary points are mapped to the set of values given by the labels of all adjacent blocks. This leads to the definition of a *spatial mapping* of type L as a total mapping $\pi : \mathbb{R}^2 \to L \cup 2^L$.

The *range* of a spatial mapping π yields the set of labels actually used in π and is denoted by $range(\pi)$. The blocks of a spatial mapping π are point sets that are mapped to the same labels. The block for a single label l (or a set S of labels) is given[6] by $f^{-1}(l)$ $(f^{-1}(S))$. The common label of a block b of π is denoted by $\pi[b]$, i.e., $\pi(b) = \{l\} \Rightarrow \pi[b] = l$. Obviously, the cardinality of block labels identifies different parts of a partition. A *region* of π is any block of π that is mapped to a single element of L, and a *border* of π is given by a block that is mapped to a set of L-values, or formally for a spatial mapping π of type L:

(i) $\rho(\pi) = \pi^{-1}(range(\pi) \cap L)$ (*regions*)
(ii) $\beta(\pi) = \pi^{-1}(range(\pi) \cap 2^L)$ (*borders*)

Now we can finally define a spatial partition by topologically constraining regions to regular open sets and by semantically constraining boundary labels to those of adjacent regions.

A *spatial partition* of type L is a spatial mapping π of type L with

(i) $\forall\ r \in \rho(\pi) : r$ is a regular open set (i.e., $r = int_T(cl_T(r))$)
(ii) $\forall\ b \in \beta(\pi) : \pi[b] = \{\pi[r] \mid r \in \rho(\pi) \wedge b \subseteq cl_T(r)\}$

The set of all spatial partitions of type L is denoted by $[L]$, i.e., $[L] \subseteq \mathbb{R}^2 \to L \cup 2^L$.

Using the representation based on α-level regions defined in the preceding subsection, we are now able to define a fuzzy region as a spatial partition. In our case $L = \Lambda_{\tilde{A}}$, i.e., the labels are formed by all possible membership values α. We have now to determine the different blocks for regions and borders. The regions of $\tilde{A}$ are given[7] by the set $\{int_T(A_{\alpha_i} -_c A_{\alpha_{i-1}}) \mid i \in \{2, \ldots, n+1\}\}$, and the borders of $\tilde{A}$ are represented by the set $\{bound_T(A_{\alpha_i} -_c A_{\alpha_{i-1}}) \mid i \in \{2, \ldots, n+1\}\}$. The object $A_{\alpha_i} -_c A_{\alpha_{i-1}}$ is a region possibly with holes. Each region is uniquely associated with an $\alpha \in \Lambda_{\tilde{A}}$, and each border has all α-labels of adjacent regions.

A *fuzzy region* $\tilde{A}$ is a spatial partition of type $\Lambda_{\tilde{A}}$ (i.e., $\tilde{A} \in [\Lambda_{\tilde{A}}]$), called an *$\alpha$-partition*.

If $\Lambda_{\tilde{A}}$ is infinite, we get an infinite spatial partition.

8 Conclusions and Future Work

This paper lays the conceptual and formal foundation for the treatment of spatial data blurred by the feature of fuzziness. It is also a contribution to bridge the gap between the entity-oriented and field-oriented view of spatial phenomena since the transitions between both views now become more and more flowing.

[6] We use the following definition of function inverse: for $f : X \to Y$ and $\forall\ y \in Y$: $f^{-1}(y) := \{x \in X \mid f(x) = y\}$. Note that f^{-1} applied to a set yields a set of sets.

[7] In the following, the operation "$-_c$" denotes the regular difference operation on regular closed sets. The operation $bound_T$ applied to a regular closed set yields its point-set topological boundary.

The paper focuses on the design of a type system for fuzzy spatial data and leads to three fuzzy spatial data types for fuzzy points, fuzzy lines, and fuzzy regions whose structure and semantics is formally defined. The characteristic feature of the design is the modeling of smoothness and continuity which is inherent to the objects themselves and to the transitions between different fuzzy objects. This is achieved by the framework of fuzzy set theory and fuzzy topology which allow partial and multiple membership and hence different membership degrees of an element in sets. Different structured views of fuzzy regions as special collections of crisp regions enable us to obtain a better understanding of their nature and to decrease their complexity.

Future work will have to deal with the formal definition of fuzzy spatial operations and predicates, with the integration of fuzzy spatial data types into query languages, and with implementation aspects leading to sophisticated data structures for the types and efficient algorithms for the operations.

References

[Alt94] D. Altman. Fuzzy Set Theoretic Approaches for Handling Imprecision in Spatial Analysis. *Int. Journal of Geographical Information Systems*, 8(3):271–289, 1994.

[BF96] P.A. Burrough and A.U. Frank, editors. *Geographic Objects with Indeterminate Boundaries*, volume 2 of *GISDATA Series*. Taylor & Francis, 1996.

[Bla84] M. Blakemore. Generalization and Error in Spatial Databases. *Cartographica*, 21, 1984.

[Bur96] P.A. Burrough. Natural Objects with Indeterminate Boundaries. *In [BF96]*, pages 3–28, 1996.

[CF96] E. Clementini and P. di Felice. An Algebraic Model for Spatial Objects with Indeterminate Boundaries. *In [BF96]*, pages 153–169, 1996.

[CG96] A.G. Cohn and N.M. Gotts. The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries. *In [BF96]*, pages 171–187, 1996.

[Cha68] C.L. Chang. Fuzzy Topological Spaces. *Journal of Mathematical Analysis and Applications*, 24:182–190, 1968.

[Dut89] S. Dutta. Qualitative Spatial Reasoning: A Semi-Quantitative Approach Using Fuzzy Logic. *1st Int. Symp. on the Design and Implementation of Large Spatial Databases (SSD'89)*, Springer-Verlag, LNCS 409:345–364, 1989.

[Dut91] S. Dutta. Topological Constraints: A Representational Framework for Approximate Spatial and Temporal Reasoning. *2nd Int. Symp. on Advances in Spatial Databases (SSD'91)*, Springer-Verlag, LNCS 525:161–180, 1991.

[ECF94] M.J. Egenhofer, E. Clementini, and P. di Felice. Topological Relations between Regions with Holes. *Int. Journal of Geographical Information Systems*, 8(2):128–142, 1994.

[ES97a] M. Erwig and M. Schneider. Partition and Conquer. *3rd Int. Conf. on Spatial Information Theory (COSIT'97)*, Springer-Verlag, LNCS 1329:389–408, 1997.

[ES97b] M. Erwig and M. Schneider. Vague Regions. *5th Int. Symp. on Advances in Spatial Databases (SSD'97)*, Springer-Verlag, LNCS 1262:298–320, 1997.

[Fin93] J.T. Finn. Use of the Average Mutual Information Index in Evaluating Classification Error and Consistency. *Int. Journal of Geographical Information Systems*, 7(4):349–366, 1993.

[Gaa64] S. Gaal. *Point Set Topology*. Academic Press, 1964.

[GS95] R.H. Güting and M. Schneider. Realm-Based Spatial Data Types: The ROSE Algebra. *VLDB Journal*, 4:100–143, 1995.

[KV91] V.J. Kollias and A. Voliotis. Fuzzy Reasoning in the Development of Geographical Information Systems. *Int. Journal of Geographical Information Systems*, 5(2):209–223, 1991.

[LAB96] P. Lagacherie, P. Andrieux, and R. Bouzigues. Fuzziness and Uncertainty of Soil Boundaries: From Reality to Coding in GIS. *In [BF96]*, pages 275–286, 1996.

[PS85] F.P. Preparata and M.I. Shamos. *Computational Geometry*. Springer Verlag, 1985.

[Sch96] M. Schneider. Modelling Spatial Objects with Undetermined Boundaries Using the Realm/ROSE Approach. *In [BF96]*, pages 141–152, 1996.

[Sch97] M. Schneider. *Spatial Data Types for Database Systems - Finite Resolution Geometry for Geographic Information Systems*, volume LNCS 1288. Springer-Verlag, Berlin Heidelberg, 1997.

[Shi93] R. Shibasaki. A Framework for Handling Geometric Data with Positional Uncertainty in a GIS Environment. *GIS: Technology and Applications*, pages 21–35, World Scientific, 1993.

[Til80] R.B. Tilove. Set Membership Classification: A Unified Approach to Geometric Intersection Problems. *IEEE Transactions on Computers*, C-29:874–883, 1980.

[Use96] E. L. Usery. A Conceptual Framework and Fuzzy Set Implementation for Geographic Features. *In [BF96]*, pages 71–85, 1996.

[Wan94] F. Wang. Towards a Natural Language User Interface: An Approach of Fuzzy Query. *Int. Journal of Geographical Information Systems*, 8(2):143–162, 1994.

[WB93] M.F. Worboys and P. Bofakos. A Canonical Model for a Class of Areal Spatial Objects. *3rd Int. Symp. on Advances in Spatial Databases (SSD'93)*, Springer-Verlag, LNCS 692:36–52, 1993.

[WHS90] F. Wang, G.B. Hall, and Subaryono. Fuzzy Information Representation and Processing in Conventional GIS Software: Database Design and Application. *Int. Journal of Geographical Information Systems*, 4(3):261–283, 1990.

[Zad65] L.A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338–353, 1965.