is that the different semantics of the uncertainty of spatial objects do not influence the representation and management of the spatial objects themselves.

This chapter is organized as follows. In Section 9.2, we propose an ontology of imperfection (taken from [17]) that is useful for understanding how imperfect observations of geographic reality cause uncertainty in spatial data. In Section 9.3, we propose various semantic interpretations of spatial objects with uncertainty. In Section 9.4, we introduce the geometric model for objects with a broad boundary, recalling the definitions for regions with broad boundary and proposing the definition of uncertain lines. Section 9.5 describes the most widely used models for topological relations (the 9-intersection [10] and the CBM [7]) applied to objects with uncertainty. In Section 9.6, we propose specific models for topological relations between uncertain lines and points. Section 9.7 uses the concept of crisping to find additional topological constraints in the case of aggregated objects, such as tessellations and networks. Section 9.8 draws short conclusions.

9.2 Ontology of imperfection

Imperfection is an essential aspect of almost all observations of natural phenomena. This is partly because many observational and measurement devices have some inherent level of imprecision and inaccuracy, but also because many representations of the natural world are inherently vague. In the next few paragraphs, we expand upon these ideas and refine our notions of imperfection through the construction of a simple ontology of imperfection (see Figure 9.1).

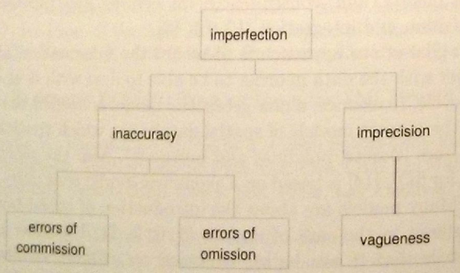

Figure 9.1. Ontology of imperfection.

Researchers use the words 'inaccuracy' and 'imprecision' in different ways. In our ontology, there is a clear distinction between inaccuracy and imprecision, and in fact these concepts are orthogonal. An *inaccurate* proposition

is one that lacks correlation with the actual state of affairs, whereas an *imprecise* proposition lacks specificity. Thus, we might propose that the time of writing this sentence is January 1st in the year 1900, at 1530 in the afternoon – fairly precise, but quite inaccurate. On the other hand, we might record the same time by merely observing that the sentence was written in the year 2000, actually accurate but imprecise.

Vagueness is that particular kind of imprecision where there are borderline cases for which it is difficult to decide whether they are covered by the concept or not. Many concepts (Bertrand Russell [15] has argued that all of language) contain vagueness. So, for example, 'tallness' is vague, because there will be people for whom is it difficult to decide whether they are tall or not. However, not all imprecise propositions are vague; For example, we could say that we are located in Europe, which is not a vague statement (assuming for the moment a crisp boundary to Europe), but is quite imprecise.

There are two types of errors that are distinguished in our ontology. The distinction here is similar to the distinction between Type I and Type II error in statistical significance testing. In Type I error, a true null hypothesis is incorrectly rejected, while in Type II error, a false null hypothesis fails to be rejected. A Type II error is only an error in the sense that an opportunity to reject the null hypothesis correctly is lost. It is not an error in the sense that an incorrect inference is made, as no conclusion is drawn when the null hypothesis is not rejected.

By analogy with Type I and II errors in statistical significance testing, we distinguish between an *error of commission* (EC) and an *error of omission* (EO). An error of commission (EC) is made when we conclude that a proposition is definitely the case when in fact it is either undetermined or definitely not the case. An error of omission (EO) is made when we do not declare a proposition to be the case when it is the case. For an example of EC, we may be using a remotely sensed image to determine land use at a location, and despite cloud cover, we wrongly assume the location to be a forest because some neighboring pixels might suggest forest. For an example of EO, we might fail to infer that we were in Switzerland because we judged some scanty evidence based on mountain scenery to be insufficient.

In summary, an ontology has been constructed that contains the following components of imperfection in observations and the data resulting from them. Inaccuracy and imprecision are proper and disjoint subconcepts of imperfection, and vagueness is a proper subconcept of imprecision. Also distinguished are the types of inaccuracy referred to as errors of omission and errors of commission. An error of commission occurs when something is declared to be the case when in fact it is not the case. An error of omission occurs when something is not declared to be the case when in fact it is the case.

Scanned by CamScanner

9.3 Spatial objects with uncertainty

In this section, we will discuss the semantic interpretations of various kinds of spatial objects affected by uncertainty. In the course of the discussion we use the concept *uncertainty* to indicate a "cognitive state induced by imperfect observations and states of knowledge".

9.3.1 Regions with broad boundary

In the case when the spatial phenomenon under observation is something as simple as a region, then the 3-valued indeterminacy of location in the region has been represented by several authors as a region with broad boundary (see Figure 9.2), also known as the egg-yolk diagram [3, 6, 9, 11].

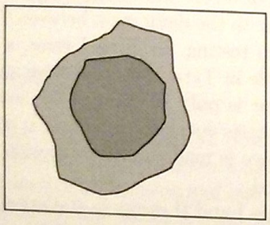

Figure 9.2. A region with a broad boundary.

Broad boundaries are a geometric model that can be used to represent uncertainty in many different situations. Therefore, a representation of a spatial phenomenon as a region with broad boundary can have several semantic interpretations (as discussed in [17]):

1. *Incomplete representation of a feature.* In vector databases, if there are missing sides of a polygon or entirely missing polygons, due to omissions in digitization or imperfect data conversion, broad boundaries can be introduced to represent the possible location of the missing lines.
2. *Conflicting representations of a feature.* In the case of cadastral data, or representations of political boundaries (in general, the case of existing data that provide approximations to some real objects), broad boundaries result from the merging of different representation of the same region.
3. *Changing representations of a dynamic spatial phenomenon.* If there are different observations of the same geographic phenomenon taken at different times, the core region may be interpreted as being the collection of locations that are in the region at all times, while the broad boundary may be interpreted as the collection of locations that are in the region at some, but not all, times.

4. *Imprecise observation of a spatial phenomenon.* At a certain resolution (i.e., a partition of the space in which locations indiscernible from the observation are grouped into the same elementary unit of the space), the elementary units impose a granularity of the underlying space. The region with broad boundary represents an imprecise observation of a region at a certain resolution.
5. *Inherently vague representations of real world spatial objects.* Most geographical objects which are not man-made artifacts or conventions fall in this category. They may be represented by fuzzy sets, if a reasonable method of quantifying the membership function for the fuzzy set is available. A region with a broad boundary is interpreted as an approximation in which the broad boundary of the region represents the part of the object where the membership function gradually decreases from 1 to 0. There are two special cases:

   a) *Inherently vague representations of real world spatial objects resulting from scale change.* These are spatial objects that only exist at coarser resolutions (small scale), but are constructed from more finely grained real world objects (large scale). Such objects may be obtained by semantic generalization, aggregating other objects of a different nature that exist at a larger scale: this is the case of urban settlements that are seen as an aggregation of other objects (such as houses, streets, subways, bridges, and so on), or woods that are made up of trees. The broad boundary of such objects represents a peripheral zone where the density of the elements composing the aggregate has an intermediate value.
   b) *Inherently vague representations of real world spatial objects resulting from variation of context.* These correspond to linguistic propositions made up of a geographic location plus a qualitative modifier. The boundary of a region as 'the south of England' is dependent on the context in which the linguistic proposition is placed and its intended use. It might be a different region if the proposition is placed in a historic, architectural, or biological context; finally, it might even depend on different opinions of individuals. Also, the qualitative modifier, like all qualitative spatial terms [8], has a different meaning depending on the granularity with which it is defined: for example, in the two domains for qualitative modifiers {south, center, north} and {south, north}, the proposition the 'south of England' would subsume a rather different underlying region.

9.3.2 Uncertain lines

Uncertainty about lines is rather more complicated to model than in the case of regions. In fact, taking a simple line (having the boundary made up of two distinct endpoints), it is not enough to consider a broad boundary instead of

Scanned by CamScanner

a crisp boundary, like in the case of regions. This would take into account a kind of uncertainty only, the one related to the endpoints of the line. We must also consider the uncertainty affecting the interior of the line, which can vary independently of the boundary, and can be modeled by a broad interior. Joining the two contributions, we define the concept of uncertain line, which takes into account both broad boundaries and broad interiors.

To understand the nature of an uncertain line, it is helpful to introduce the concept of a crisping: a crisping of an uncertain line is any line with endpoints and interior that are strictly contained in the uncertain line. Also, un uncertain line can be seen as the set of all possible crispings. Uncertainty affecting an endpoint can be represented by a region (Figure 9.3(a)). Such a region is the set of points such that they can be an endpoint of a crisping. We assume that for each endpoint there is a simple region and call the union of those regions the broad boundary of the uncertain line. This definition of broad boundary of a line has the same semantics of broad boundaries of regions: also for the latter ones, broad boundaries are the set of points that could belong to the boundary of the region's crispings. A problem is that uncertainty on boundary implies some uncertainty on the interior as well, as it is illustrated in Figure 9.3(a). In fact, the crisping of a line starting from an endpoint differs from other crispings also for a piece of the interior.

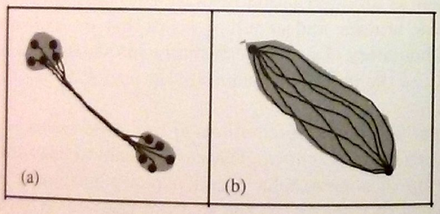

Figure 9.3. Broad boundary (a) and broad interior (b) of a simple line.

The second kind is uncertainty affecting the interior, when endpoints are crisp. Such a kind of uncertainty can be represented by the region of all possible crispings connecting the endpoints, as illustrated in Figure 9.3(b). In this region, that we call broad interior, we may find all points that could belong to the interior of a crisping. The concept of broad interior has not a direct correspondence in the case of regions, for which the uncertainty on boundaries takes into account all kinds of uncertainty.

The concept of uncertain line encompasses the two types of uncertainty (see Figure 9.4(a)). We can draw some conclusions about the topology of regions representing broad boundaries and broad interiors. If a point could be part of the boundary of a line's crisping, it means that a neighborhood of that point contains points of the interior of a crisping. This proves that a point of

the broad boundary cannot be outside the broad interior. Therefore, the broad boundary is inside the broad interior. A broad interior is always present, even if the uncertainty is only about endpoints. Therefore, the region representing the broad interior is not necessarily a regularly closed set, because in general it may have two-dimensional and one-dimensional parts, as illustrated in Figure 9.4(b). However, it must be connected to be able to contain all possible crispings.

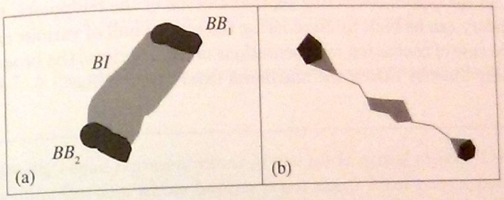

Figure 9.4. An uncertain line (a) with its broad interior (BI) and broad boundary (BB_1 and BB_2). An uncertain line whose broad interior has two-dimensional and one-dimensional parts (b).

The following are the possible semantic interpretations of uncertain lines:

1. *Incomplete representation of a line feature.* An example of this case is the situation sometimes arising in vector databases, where in a vector representation of a line feature there might be missing lines or segments of lines, due to omissions in digitization or imperfect data conversion. The broad interior (shaded area in Figure 9.5) can be introduced to represent the possible location of the missing parts. In this case, the phenomena could be represented by crisp concepts, but the observations we have at disposal are not sufficient to be fully specific.

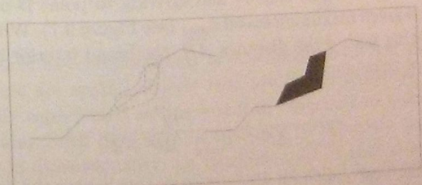

Figure 9.5. A broad interior to model an incomplete representation.

2. *Conflicting representations of a line feature.* This might arise in the case of cadastral data, or representations of political boundaries: in general, the case of existing data that provide approximations to some real objects.

Scanned by CamScanner

There might be both errors of commission and omission depending on which processes took place to obtain the representations. Different logics can be applied during the amalgamation of two resolutions. In the case of ignorance of the origin of the data, the most pessimistic view would need to be adopted (both types of errors in the two resolutions). Figure 9.6 shows an example of this, where a line with broad boundary results from the merging of different positions of the same line. It can be the case of conflicting representations both of the endpoints and of interior of the line. In the case of conflicting representations of the endpoints, the broad boundary can be built by considering the convex hull of various endpoints. In the case of conflicting representations of the interiors, the broad interior is easily built by taking the maximum extent of the lines.

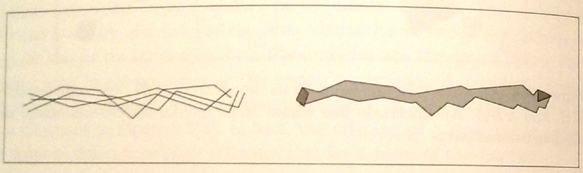

Figure 9.6. An uncertain line to model conflicting representations of endpoints and interior.

3. *Changing representations of a dynamic spatial phenomenon.* Variability over time applies when we have different observations of the same geographic phenomenon taken at different times. Given a line that moves its position over times t_1 and t_2, the method of building the broad boundary and the broad interior of the corresponding uncertain line is the same as in previous case. As another example of this category, a line with a broad interior can represent the set of all possible trajectories of a moving agent starting from point A at time t_1 and arriving to point B at time t_2 proceeding at a given maximum speed $v_{\max}$ (see Figure 9.7). With $\Delta t = t_2 - t_1$ and indicating with d the distance $\overline{AB}$, the broad interior is bounded by an ellipse of equation:

$$\frac{4x^2}{v_{\max}^2 \cdot \Delta t^2} + \frac{4y^2}{v_{\max}^2 \cdot \Delta t^2 - d^2} = 1.$$

The equation above holds for a reference system having the points A and B on the x axis and equidistant from the origin. It is obtained by considering that the maximum distance that can be travelled by the agent is $v_{\max} \cdot \Delta t$. If the trajectories are straight segments, the moving agent can reach all the points P lying on the ellipse, which, by definition, are the points having constant the sum of distances $\overline{AP} + \overline{PB}$. If the trajectories are curvilinear, they must lie inside the ellipse.

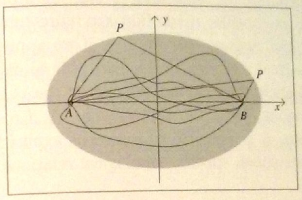

Figure 9.7. A line with broad interior to represent possible trajectories of a moving agent.

4. *Inherently vague representations of real world spatial objects.* This case includes line features whose interior is not sharp. Most geographical objects which are not man-made artifacts or conventions fall in this category: for the case of line features, we include in this category rivers, coastal lines, animal paths in the forest, etc. In Figure 9.8, a broad interior is helpful for the representation of a river (and a broad boundary for its endpoint can represent its mouth).

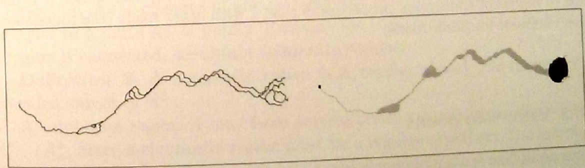

Figure 9.8. An uncertain line to model an inherently vague representation of a real world spatial object (for example, a river).

5. *Line features as representations of two-dimensional real world spatial objects.* These are spatial objects that are two-dimensional in reality, but are usually represented as line features, especially at coarser resolutions (small scale). Such lines may be obtained by semantic generalization by a change in dimensionality: for example a street that is represented as a two-dimensional parcel at a large scale and as a line feature at a small scale. These objects have typically a prevalent dimension (roads, highways, railways, utilities lines etc.) and their connection properties are more important than their size and shape: this is why they are usually represented as line features. The proposed model for uncertain lines can be useful in this case for maintaining the two-dimensional representation of the line especially at intermediate scales where the width of the line is still relevant

and, at the same time, for remarking connection properties. To illustrate the concept, let us consider roads that are represented at a certain scale by two-dimensional regions (Figure 9.9): if we apply the models for regions, the topological relations that we are able to distinguish are region-oriented and as such they cannot emphasize the characteristics of lines (connection, crossing, etc.). On the other hand, if roads are represented by uncertain lines, it is possible to represent explicitly the endpoints of the lines and distinguish specific topological relations between two roads (see Section 9.6).

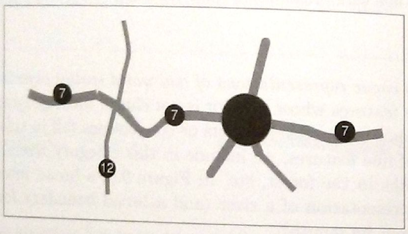

Figure 9.9. Uncertain lines as representation of roads: using this model, it is possible to represent connection properties of road 7 with other roads (the roundabout is an endpoint for other roads).

9.3.3 Uncertain points

For a point-like spatial phenomenon, the main source of uncertainty is related to the indetermination of its position in space. We call uncertain point the region of all possible positions. Un uncertain point may be represented as an open set that we may assume to be a simply connected region. This definition will allow us to easily reapply the same definitions for topological relations that we had for points, as we will see in Section 9.6.

The semantic interpretations of an uncertain point are the following:

1. *Endpoints of uncertain lines.* Often, the endpoints of uncertain lines can be seen as independent geographic features, like in the previous examples concerning a river, where an endpoint is the mouth of the river, and roads, where endpoints may represent roundabouts and crossroads.
2. *1-dimensional landmarks.* At a small scale, many geographic features are represented by a point (a lighthouse, an archeological site, etc.). The same features at a large scale are represented by regions. At a an intermediate scale, it might be the case that a point-based representation is too imprecise to represent the feature, while a region-based representation looses the properties of being point-like. A representation based on an uncertain point could be a solution, because it has an extension to take into account imprecision and topological relations for point-like features can be applied to it (see Section 9.6).
3. *Positional uncertainty.* Many point-based geographic features can be known only with some positional uncertainty due to measurement errors (e.g., the peak of a mountain, a wreck on the bottom of sea).

9.4 The geometric model of objects with uncertainty

In this section, we give the geometric definitions of various spatial objects affected by uncertainty. We start with regions with a broad boundary that were originally defined in [3, 6]. As preliminary notions, we first define crisp regions and composite regions:

Definition 1. A *region* is a regular closed two-dimensional bounded subset of $\mathbb{R}^2$ with a connected interior.

Definition 1 does not exclude the presence of holes. Therefore, the exterior of a region A may be separated in several components: there exist an outer exterior (unbounded set) denoted by A_0^- and $n \geq 0$ inner exteriors (bounded sets) denoted by $A_1^- \dots A_n^-$. Since a region must be a regular closed set, the intersection of the closures of any two different exterior components is empty or equal to a finite set of points. Relaxing the constraint that the interior of a region is connected, we obtain composite regions:

Definition 2. A *composite region* is a regular closed two-dimensional bounded subset of $\Re^2$.

A composite region A may have several interior components denoted by $A_1^\circ \dots A_n^\circ$. Since a composite region must be a regular closed set, the intersection of the closures of any two different interior components is empty or equal to a finite set of points.

Below, we give the definition of regions with a *broad boundary*, which differ from crisp ones with regard to the boundary definition. For a region with a broad boundary, we can define an *inner boundary* and an *outer boundary*, where the inner boundary is surrounded by the outer boundary. The closed annular region comprised between the inner and outer boundary is the broad boundary of the original region.

Definition 3. A *region with a broad boundary* A is made up of two regions A_1 and A_2, with $A_1 \subseteq A_2$, where ∂A_1 is the *inner boundary* of A and ∂A_2 is the *outer boundary* of A.

Definition 4. The *broad boundary* ΔA of a region with a broad boundary A is the closed subset comprised between the inner boundary and the outer boundary of A, i.e., $\Delta A = \overline{A_2 - A_1}$, or equivalently $\Delta A = A_2 - A_1^\circ$.

Definition 5. Interior, closure, and exterior of a region with a broad boundary A are defined as $A^\circ = A_2 - \Delta A$, $\overline{A} = A^\circ \cup \Delta A$, $A^- = \mathbb{R}^2 - \overline{A}$, respectively.

Scanned by CamScanner

The interior and exterior of a region with a broad boundary are open sets, while the broad boundary is a closed set. Notice that both the exterior and broad boundary of a region A may have several components, because regions A_1 and A_2 may have holes. Figure 9.10 illustrates some cases that can arise: case (a) is a region without holes; case (b) is a region with two holes; case (c) is a region where A_1 has one hole and A_2 has two holes; case (d) is a region where A_1 has two holes and A_2 has one hole.

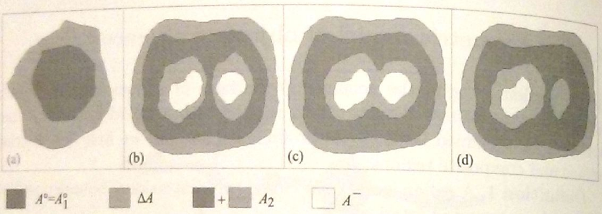

Figure 9.10. Regions with a broad boundary.

Finally, we consider the most general case of composite regions with a broad boundary:

Definition 6. A *composite region with a broad boundary* A is made up of two composite regions A_1 and A_2, with $A_1 \subseteq A_2$, where ∂A_1 is the *inner boundary* of A and ∂A_2 is the *outer boundary* of A.

Definition 7. The *broad boundary* ΔA of a composite region with a broad boundary A is the closed subset comprised between the inner boundary and the outer boundary of A, i.e., $\Delta A = \overline{A_2 - A_1}$, or equivalently $\Delta A = A_2 - A_1^\circ$.

Definition 8. Interior, closure, and exterior of a composite region with a broad boundary A are defined as $A^\circ = A_2 - \Delta A$, $\overline{A} = A^\circ \cup \Delta A$, $A^- = \mathbb{R}^2 - \overline{A}$, respectively.

Figure 9.11 illustrates some configurations of composite regions: case (a) is a region with two components; case (b) is a region where A_1 has two components and A_2 has one component; case (c) is a region where A_1 has one component and A_2 has two components.

The initial definitions of lines with uncertainty were given in [5], where authors distinguished two kinds of broad boundaries, modeling either the position of the line or the position of its endpoints. The first kind of broad boundary was an area surrounding the whole line, while the second kind was made up of an area for each endpoint. These definitions are not satisfactory with respect to the topological relations that can be defined over them. A new definition is needed, which takes into account both the uncertainty on endpoints and on the interior of the line. We first define broad interiors and broad boundaries and after that we define the concept of *uncertain line*. A

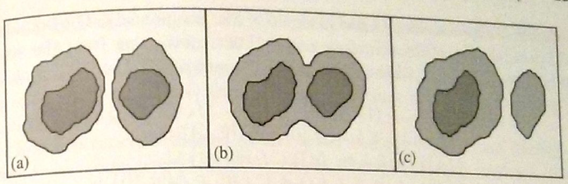

Figure 9.11. Composite regions with a broad boundary.

first attempt to give such definitions was given in [2]. As preliminary notions, we recall the definitions of lines without uncertainty.

Definition 9. A *simple line* is a closed (non-empty) one-dimensional point-set, L, defined as the image of a continuous mapping $f: [0,1] \to \mathbb{R}^2$, such that $f(t_i) \neq f(t_j)$, $\forall t_i, t_j \in [0,1]$, $t_i \neq t_j$.

The mappings of 0 and 1 through f are the two endpoints of L, these two points made up the boundary of a simple line in the plane; i.e., $\partial L = \{f(0), f(1)\}$. By relaxing the constraint of no self-intersections in the interior, we get the notion of *line with self-intersections*. Figure 9.12 shows a simple line and four lines with self-intersections. The number of endpoints for a line with self-intersections can be either 2, 1, or 0. In detail, the number of endpoints of L is equal to:

- 2 if $f(0) \notin f((0,1]) \wedge f(1) \notin f([0,1))$; $\partial L = \{f(0), f(1)\}$;
- 1 if $f(0) \in f((0,1)) \wedge f(1) \notin f((0,1))$; $\partial L = \{f(1)\}$;
  or if $f(0) \notin f((0,1)) \wedge f(1) \in f((0,1))$; $\partial L = \{f(0)\}$;
- 0 if $(f(0) = f(1)) \vee (f(0) \in f((0,1]) \wedge f(1) \in f([0,1)))$; $\partial L = \emptyset$.

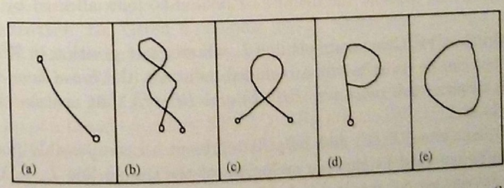

Figure 9.12. A simple line (a) and four lines with self-intersections (b-e).

Definition 10. Let $f_1, f_2, \ldots f_n$ be continuous mappings from the interval [0, 1] to the plane. We call *complex line* any closed (non-empty) one-dimensional point-set, L, defined as the union of the image of the functions $f_1, f_2, \ldots f_n$:

$$f_1([0,1]) \cup f_2([0,1]) \cup \ldots \cup f_n([0,1]).$$

Scanned by CamScanner