

Chapter I

Spatial Data Warehouse Modelling

Maria Luisa Damiani, DICO - University of Milan, Italy

Stefano Spaccapietra, Ecole Polytechnique Fédérale, Switzerland

Abstract

This chapter is concerned with multidimensional data models for spatial data warehouses. Over the last few years different approaches have been proposed in the literature for modelling multidimensional data with geometric extent. Nevertheless, the definition of a comprehensive and formal data model is still a major research issue. The main contributions of the chapter are twofold: First, it draws a picture of the research area; second it introduces a novel spatial multidimensional data model for spatial objects with geometry (MuSD--multigranular spatial data warehouse). MuSD complies with current standards for spatial data modelling, augmented by data warehousing concepts such as spatial fact, spatial dimension and spatial measure. The novelty of the model is the representation of spatial measures at multiple levels of geometric granularity. Besides the representation concepts, the model includes a set of OLAP operators supporting the navigation across dimension and measure levels.

Introduction

A topic that over recent years that has received growing attention from both academy and industry concerns the integration of spatial data management with multidimensional data analysis techniques. We refer to this technology as spatial data warehousing, and consider a spatial data warehouse to be a multidimensional database of spatial data. Following common practice, we use here the term spatial in the geographical sense, i.e., to denote data that includes the description of how objects and phenomena are located on the Earth. A large variety of data may be considered to be spatial, including: data for land use and socioeconomic analysis; digital imagery and geo-sensor data; location-based data acquired through GPS or other positioning devices; environmental phenomena. Such data are collected and possibly marketed by organizations such as public administrations, utilities and other private companies, environmental research centres and spatial data infrastructures. Spatial data warehousing has been recognized as a key technology in enabling the interactive analysis of spatial data sets for decision-making support (Rivest et al., 2001; Han et al., 2002). Application domains in which the technology can play an important role are, for example, those dealing with complex and worldwide phenomena such as homeland security, environmental monitoring and health safeguards. These applications pose challenging requirements for integration and usage of spatial data of different kinds, coverage and resolution, for which the spatial data warehouse technology may be extremely helpful.

Origins

Spatial data warehousing results from the confluence of two technologies, spatial data handling and multidimensional data analysis, respectively. The former technology is mainly provided by two kinds of systems: spatial database management systems (DBMS) and geographical information systems (GIS). Spatial DBMS extend the functionalities of conventional data management systems to support the storage, efficient retrieval and manipulation of spatial data (Rigaux et al., 2002). Examples of commercial DBMS systems are Oracle Spatial and IBM DB2 Spatial Extender. A GIS, on the other hand, is a composite computer based information system consisting of an integrated set of programs, possibly including or interacting with a spatial DBMS, which enables the capturing, modelling, analysis and visualization of spatial data (Longley, et al., 2001). Unlike a spatial DBMS, a GIS is meant to be directly usable by an end-user. Examples of commercial systems are ESRI ArcGIS and Intergraph Geomedia. The technology of spatial data handling has made significant progress in the last decade, fostered by the standardization initiatives promoted by OGC

(Open Geospatial Consortium) and ISO/TC211, as well as by the increased availability of off-the-shelf geographical data sets that have broadened the spectrum of spatially-aware applications. Conversely, multidimensional data analysis has become the leading technology for decision making in the business area. Data are stored in a multidimensional array (cube or hypercube) (Kimball, 1996; Chaudhuri & Dayla, 1997; Vassiliadis & Sellis, 1999). The elements of the cube constitute the facts (or cells) and are defined by measures and dimensions. Typically, a measure denotes a quantitative variable in a given domain. For example, in the marketing domain, one kind of measure is sales amount. A dimension is a structural attribute characterizing a measure. For the marketing example, dimensions of sales may be: time, location and product. Under these example assumptions, a cell stores the amount of sales for a given product in a given region and over a given period of time. Moreover, each dimension is organized in a hierarchy of dimension levels, each level corresponding to a different granularity for the dimension. For example, *year* is one level of the *time* dimension, while the sequence *day*, *month*, *year* defines a simple hierarchy of increasing granularity for the time dimension. The basic operations for online analysis (OLAP operators) that can be performed over data cubes are: *roll-up*, which moves up along one or more dimensions towards more aggregated data (e.g., moving from monthly sales amounts to yearly sales amounts); *drill-down*, which moves down dimensions towards more detailed, disaggregated data and *slice-and-dice*, which performs a selection and projection operation on a cube. The integration of these two technologies, spatial data handling and multidimensional analysis, responds to multiple application needs. In business data warehouses, the spatial dimension is increasingly considered of strategic relevance for the analysis of enterprise data. Likewise, in engineering and scientific applications, huge amounts of measures, typically related to environmental phenomena, are collected through sensors, installed on ground or satellites, and continuously generating data to be stored in data warehouses for subsequent analysis.

Spatial Multidimensional Models

A data warehouse (DW) is the result of a complex process entailing the integration of huge amounts of heterogeneous data, their organization into denormalized data structures and eventually their loading into a database for use through online analysis techniques. In a DW, data are organized and manipulated in accordance with the concepts and operators provided by a multidimensional data model. Multidimensional data models have been widely investigated for conventional, non-spatial data. Commercial systems based on these models are marketed. By contrast, research on spatially aware DWs (SDWs) is a step

behind. The reasons are diverse: The spatial context is peculiar and complex, requiring specialized techniques for data representation and processing; the technology for spatial data management has reached maturity only in recent times with the development of SQL3-based implementations of OGC standards; finally, SDWs still lack a market comparable in size with the business sector that is pushing the development of the technology. As a result, the definition of spatial multidimensional data models (SMDs) is still a challenging research issue.

A SMD model can be specified at conceptual and logical levels. Unlike the logical model, the specification at the conceptual level is independent of the technology used for the management of spatial data. Therefore, since the representation is not constrained by the implementation platform, the conceptual specification, that is the view we adopt in this work, is more flexible, although not immediately operational.

The conceptual specification of an SMD model entails the definition of two basic components: a set of representation constructs, and an algebra of spatial OLAP (SOLAP) operators, supporting data analysis and navigation across the representation structures of the model. The representation constructs account for the specificity of the spatial nature of data. In this work we focus on one of the peculiarities of spatial data, that is the availability of spatial data at different levels of granularity. Since the granularity concerns not only the semantics but also the geometric aspects of the data, the location of objects can have different geometric representations. For example, representing the location of an accident at different scales may lead to associating different geometries to the same accident.

To allow a more flexible representation of spatial data at different geometric granularity, we propose a SDM model in which not only dimensions are organized in levels of detail but also the spatial measures. For that purpose we introduce the concept of *multi-level spatial measure*.

The proposed model is named MuSD (multigranular spatial data warehouse). It is based on the notions of *spatial fact*, *spatial dimension* and *multi-level spatial measure*. A spatial fact may be defined as a fact describing an event that occurred on the Earth in a position that is relevant to know and analyze. Spatial facts are, for instance, road accidents. Spatial dimensions and measures represent properties of facts that have a geometric meaning; in particular, the spatial measure represents the location in which the fact occurred. A multi-level spatial measure is a measure that is represented by multiple geometries at different levels of detail. A measure of this kind is, for example, the location of an accident: Depending on the application requirements, an accident may be represented by a point along a road, a road segment or the whole road, possibly at different cartographic scales. Spatial measures and dimensions are uniformly represented in terms of the standard spatial objects defined by the Open

Geospatial Consortium. Besides the representation constructs, the model includes a set of SOLAP operators to navigate not only through the dimensional levels but also through the levels of the spatial measures.

The chapter is structured in the following sections: the next section, *Background Knowledge*, introduces a few basic concepts underlying spatial data representation; the subsequent section, *State of the Art on Spatial Multidimensional Models*, surveys the literature on SDM models; the proposed spatial multidimensional data model is presented in the following section; and research opportunities and some concluding remarks are discussed in the two conclusive sections.

Background Knowledge

The real world is populated by different kinds of objects, such as roads, buildings, administrative boundaries, moving cars and air pollution phenomena. Some of these objects are tangible, like buildings, others, like administrative boundaries, are not. Moreover, some of them have identifiable shapes with well-defined boundaries, like land parcels; others do not have a crisp and fixed shape, like air pollution. Furthermore, in some cases the position of objects, e.g., buildings, does not change in time; in other cases it changes more or less frequently, as in the case of moving cars. To account for the multiform nature of spatial data, a variety of data models for the digital representation of spatial data are needed. In this section, we present an overview of a few basic concepts of spatial data representation used throughout the chapter.

The Nature of Spatial Data

Spatial data describe properties of phenomena occurring in the world. The prime property of such phenomena is that they occupy a position. In broad terms, a position is the description of a location on the Earth. The common way of describing such a position is through the coordinates of a coordinate reference system.

The real world is populated by phenomena that fall into two broad conceptual categories: entities and continuous fields (Longley et al., 2001). Entities are distinguishable elements occupying a precise position on the Earth and normally having a well-defined boundary. Examples of entities are rivers, roads and buildings. By contrast, fields are variables having a single value that varies within a bounded space. An example of field is the temperature, or the distribution, of a polluting substance in an area. Field data can be directly obtained from sensors,

for example installed on satellites, or obtained by interpolation from sample sets of observations.

The standard name adopted for the digital representation of abstractions of real world phenomena is that of *feature* (OGC, 2001; OGC, 2003). The feature is the basic representation construct defined in the reference spatial data model developed by the Open Geospatial Consortium and endorsed by ISO/TC211. As we will see, we will use the concept of feature to uniformly represent all the spatial components in our model. Features are spatial when they are associated with locations on the Earth; otherwise they are non-spatial. Features have a distinguishing name and have a set of attributes. Moreover, features may be defined at instance and type level: *Feature instances* represent single phenomena; *feature types* describe the intensional meaning of features having a common set of attributes. Spatial features are further specialized to represent different kinds of spatial data. In the OGC terminology, *coverages* are the spatial features that represent continuous fields and consist of discrete functions taking values over space partitions. Space partitioning results from either the subdivision of space in a set of regular units or cells (*raster* data model) or the subdivision of space in irregular units such as triangles (*tin* data model). The discrete function assigns each portion of a bounded space a value.

In our model, we specifically consider *simple spatial features*. Simple spatial features (“features” hereinafter) have one or more attributes of geometric type, where the geometric type is one of the types defined by OGC, such as point, line and polygon. One of these attributes denotes the position of the entity. For example, the position of the state Italy may be described by a multipolygon, i.e., a set of disjoint polygons (to account for islands), with holes (to account for the Vatican State and San Marino). A simple feature is very close to the concept of entity or object as used by the database community. It should be noticed, however, that besides a semantic and geometric characterization, a feature type is also assigned a coordinate reference system, which is specific for the feature type and that defines the space in which the instances of the feature type are embedded.

More complex features may be defined specifying the topological relationships relating a set of features. Topology deals with the geometric properties that remain invariant when space is elastically deformed. Within the context of geographical information, topology is commonly used to describe, for example, connectivity and adjacency relationships between spatial elements. For example, a road network, consisting of a set of interconnected roads, may be described through a graph of nodes and edges: Edges are the topological objects representing road segments whereas nodes account for road junctions and road endpoints.

To summarize, spatial data have a complex nature. Depending on the application requirements and the characteristics of the real world phenomena, different

spatial data models can be adopted for the representation of geometric and topological properties of spatial entities and continuous fields.

State of the Art on Spatial Multidimensional Models

Research on spatial multidimensional data models is relatively recent. Since the pioneering work of Han et al., (1998), several models have been proposed in the literature aiming at extending the classical multidimensional data model with spatial concepts. However, despite the complexity of spatial data, current spatial data warehouses typically contain objects with simple geometric extent.

Moreover, while an SMD model is assumed to consist of a set of representation concepts and an algebra of SOLAP operators for data navigation and aggregation, approaches proposed in the literature often privilege only one of the two aspects, rarely both. Further, whilst early data models are defined at the logical level and are based on the relational data model, in particular on the star model, more recent developments, especially carried out by the database research community, focus on conceptual aspects. We also observe that the modelling of geometric granularities in terms of multi-level spatial measures, which we propose in our model, is a novel theme.

Often, existing approaches do not rely on standard data models for the representation of spatial aspects. The spatiality of facts is commonly represented through a geometric element, while in our approach, as we will see, it is an OGC spatial feature, i.e., an object that has a semantic value in addition to its spatial characterization.

A related research issue that is gaining increased interest in recent years, and that is relevant for the development of comprehensive SDW data models, concerns the specification and efficient implementation of the operators for spatial aggregation.

Literature Review

The first, and perhaps the most significant, model proposed so far has been developed by Han et al. (1998). This model introduced the concepts of spatial dimension and spatial measure. Spatial dimensions describe properties of facts that also have a geometric characterization. Spatial dimensions, as conventional dimensions, are defined at different levels of granularity. Conversely, a spatial measure is defined as “a measure that contains a collection of pointers to spatial

objects,” where spatial objects are geometric elements, such as polygons. Therefore, a spatial measure does not have a semantic characterization, it is just a set of geometries. To illustrate these concepts, the authors consider a SDW about weather data. The example SDW has three thematic dimensions: {temperature, precipitation, time}; one spatial dimension: {region}; and three measures: {region_map, area, count}. While area and count are numeric measures, region_map is a spatial measure denoting a set of polygons. The proposed model is specified at the logical level, in particular in terms of a star schema, and does not include an algebra of OLAP operators. Instead, the authors develop a technique for the efficient computation of spatial aggregations, like the merge of polygons. Since the spatial aggregation operations are assumed to be distributive, aggregations may be partially computed on disjoint subsets of data. By pre-computing the spatial aggregation of different subsets of data, the processing time can be reduced.

Rivest et al. (2001) extend the definition of spatial measures given in the previous approach to account for spatial measures that are computed by metric or topological operators. Further, the authors emphasize the need for more advanced querying capabilities to provide end users with topological and metric operators. The need to account for topological relationships has been more concretely addressed by Marchant et al. (2004), who define a specific type of dimension implementing spatio-temporal topological operators at different levels of detail. In such a way, facts may be partitioned not only based on dimension values but also on the existing topological relationships.

Shekhar et al. (2001) propose a *map cube operator*, extending the concepts of data cube and aggregation to spatial data. Further, the authors introduce a classification and examples of different types of spatial measures, e.g., spatial distributive, algebraic and holistic functions.

GeoDWFrame (Fidalgo et al., 2004) is a recently proposed model based on the star schema. The conceptual framework, however, does not include the notion of spatial measure, while dimensions are classified in a rather complex way.

Pederson and Tryfona (2001) are the first to introduce a formal definition of an SMD model at the conceptual level. The model only accounts for spatial measures whilst dimensions are only non-spatial. The spatial measure is a collection of geometries, as in Han et al. (1998), and in particular of polygonal elements. The authors develop a pre-aggregation technique to reduce the processing time of the operations of merge and intersection of polygons. The formalization approach is valuable but, because of the limited number of operations and types of spatial objects that are taken into account, the model has limited functionality and expressiveness.

Jensen et al. (2002) address an important requirement of spatial applications. In particular, the authors propose a conceptual model that allows the definition of

dimensions whose levels are related by a partial containment relationship. An example of partial containment is the relationship between a roadway and the district it crosses. A degree of containment is attributed to the relationship. For example, a roadway may be defined as partially contained at degree 0.5 into a district. An algebra for the extended data model is also defined. To our knowledge, the model has been the first to deal with uncertainty in data warehouses, which is a relevant issue in real applications.

Malinowski and Zimanyi (2004) present a different approach to conceptual modelling. Their SMD model is based on the Entity Relationship modelling paradigm. The basic representation constructs are those of *fact relationship* and *dimension*. A *dimension* contains one or several related levels consisting of entity types possibly having an attribute of geometric type. The *fact relationship* represents an n-ary relationship existing among the dimension levels. The attributes of the *fact relationship* constitute the measures. In particular, a spatial measure is a measure that is represented by a geometry or a function computing a geometric property, such as the length or surface of an element. The spatial aspects of the model are expressed in terms of the MADS spatio-temporal conceptual model (Parent et al., 1998). An interesting concept of the SMD model is that of *spatial fact relationship*, which models a spatial relationship between two or more spatial dimensions, such as that of spatial containment. However, the model focuses on the representation constructs and does not specify a SOLAP algebra.

A different, though related, issue concerns the operations of *spatial aggregation*. Spatial aggregation operations summarize the geometric properties of objects, and as such constitute the distinguishing aspect of SDW. Nevertheless, despite the relevance of the subject, a standard set of operators (as, for example, the operators Avg, Min, Max in SQL) has not been defined yet. A first comprehensive classification and formalization of spatio-temporal aggregate functions is presented in Lopez & Snodgrass (2005). The operation of aggregation is defined as a function that is applied to a collection of tuples and returns a single value. The authors distinguish three kinds of methods for generating the set of tuples, known as *group composition*, *partition composition* and *sliding window composition*. They provide a formal definition of aggregation for conventional, temporal and spatial data based on this distinction. In addition to the conceptual aspects of spatial aggregation, another major issue regards the development of methods for the efficient computation of these kinds of operations to manage high volumes of spatial data. In particular, techniques are developed based on the combined use of specialized indexes, materialization of aggregate measures and computational geometry algorithms, especially to support the aggregation of dynamically computed sets of spatial objects (Papadias, et. al., 2001; Rao, et. al., 2003; Zhang & Tsotras, 2005).

A Multigranular Spatial Data Warehouse Model: MuSD

Despite the numerous proposals of data models for SDW defined at the logical, and more recently, conceptual level presented in the previous section, and despite the increasing number of data warehousing applications (see, e.g., Bedard et al., 2003; Scotch & Parmantoea, 2005), the definition of a comprehensive and formal data model is still a major research issue.

In this work we focus on the definition of a formal model based on the concept of spatial measures at multiple levels of geometric granularity.

One of the distinguishing aspects of multidimensional data models is the capability of dealing with data at different levels of detail or granularity. Typically, in a data warehouse the notion of granularity is conveyed through the notion of dimensional hierarchy. For example, the dimension *administrative units* may be represented at different decreasing levels of detail: at the most detailed level as municipalities, next as regions and then as states. Note, however, that unlike dimensions, measures are assigned a unique granularity. For example, the granularity of sales may be homogeneously expressed in euros.

In SDW, the assumption that spatial measures have a unique level of granularity seems to be too restrictive. In fact, spatial data are very often available at multiple granularities, since data are collected by different organizations for different purposes. Moreover, the granularity not only regards the semantics (semantic granularity) but also the geometric aspects (spatial granularity) (Spaccapietra et al., 2000; Fonseca et al., 2002). For example, the location of an accident may be modelled as a measure, yet represented at different scales and thus have varying geometric representations.

To represent measures at varying spatial granularities, alternative strategies can be prospected: A simple approach is to define a number of spatial measures, one for each level of spatial granularity. However, this solution is not conceptually adequate because it does not represent the hierarchical relation among the various spatial representations.

In the model we propose, named MuSD, we introduce the notion of *multi-level spatial measure*, which is a spatial measure that is defined at multiple levels of granularity, in the same way as dimensions. The introduction of this new concept raises a number of interesting issues. The first one concerns the modelling of the spatial properties. To provide a homogeneous representation of the spatial properties across multiple levels, both spatial measures and dimensions are represented in terms of OGC features. Therefore, the locations of facts are denoted by feature identifiers. For example, a feature, say p1, of type *road*

accident, may represent the location of an *accident*. Note that in this way we can refer to spatial objects in a simple way using names, in much the same way Han et al. (1998) do using pointers. The difference is in the level of abstraction and, moreover, in the fact that a feature is not simply a geometry but an entity with a semantic characterization.

Another issue concerns the representation of the features resulting from aggregation operations. To represent such features at different granularities, the model is supposed to include a set of operators that are able to dynamically decrease the spatial granularity of spatial measures. We call these operators *coarsening operators*. With this term we indicate a variety of operators that, although developed in different contexts, share the common goal of representing less precisely the geometry of an object. Examples include the operators for cartographic generalization proposed in Camossi et al. (2003) as well the operators generating imprecise geometries out of more precise representations (*fuzzyfying operators*).

In summary, the MuSD model has the following characteristics:

- It is based on the usual constructs of (spatial) measures and (spatial) dimensions. Notice that the spatiality of a measure is a necessary condition for the DW to be spatial, while the spatiality of dimensions is optional;
- A spatial measure represents the location of a fact at multiple levels of spatial granularity;
- Spatial dimension and spatial measures are represented in terms of OGC features;
- Spatial measures at different spatial granularity can be dynamically computed by applying a set of coarsening operators; and
- An algebra of SOLAP operators is defined to enable user navigation and data analysis.

Hereinafter, we first introduce the representation concepts of the MuSD model and then the SOLAP operators.

Representation Concepts in MuSD

The basic notion of the model is that of *spatial fact*. A spatial fact is defined as a fact that has occurred in a location. Properties of spatial facts are described in terms of measures and dimensions which, depending on the application, may have a spatial meaning.

A *dimension* is composed of *levels*. The set of levels is partially ordered; more specifically, it constitutes a lattice. Levels are assigned values belonging to *domains*. If the domain of a level consists of features, the level is *spatial*; otherwise it is *non-spatial*. A *spatial measure*, as a dimension, is composed of levels representing different granularities for the measure and forming a lattice. Since in common practice the notion of granularity seems not to be of particular concern for conventional and numeric measures, non-spatial measures are defined at a unique level. Further, as the spatial measure represents the location of the fact, it seems reasonable and not significantly restrictive to assume the spatial measure to be unique in the SDW.

As Jensen et al. (2002), we base the model on the distinction between the intensional and extensional representations, which we respectively call *schema* and *cube*. The schema specifies the structure, thus the set of dimensions and measures that compose the SDW; the cube describes a set of facts along the properties specified in the schema.

To illustrate the concepts of the model, we use as a running example the case of an SDW of road accidents. The *accidents* constitute the spatial facts. The properties of the accidents are modelled as follows: The number of *victims* and the *position* along the road constitute the measures of the SDW. In particular, the position of the accident is a spatial measure. The *date* and the *administrative unit* in which the accident occurred constitute the dimensions.

Before detailing the representation constructs, we need to define the spatial data model which is used for representing the spatial concepts of the model.

The Spatial Data Model

For the representation of the spatial components, we adopt a spatial data model based on the OGC simple features model. We adopt this model because it is widely deployed in commercial spatial DBMS and GIS. Although a more advanced spatial data model has been proposed (OGC, 2003), we do not lose in generality by adopting the simple feature model. Features (simple) are identified by names. Milan, Lake Michigan and the car number AZ213JW are examples of features. In particular, we consider as spatial features entities that can be mapped onto locations in the given space (for example, Milan and Lake Michigan). The location of a feature is represented through a *geometry*. The geometry of a spatial feature may be of type point, line or polygon, or recursively be a collection of disjoint geometries. Features have an application-dependent semantics that are expressed through the concept of *feature type*. Road, Town, Lake and Car are examples of feature types. The *extension* of a feature type, *ft*, is a set of semantically homogeneous features. As remarked in the previous

section, since features are identified by unique names, we represent spatial objects in terms of feature identifiers. Such identifiers are different from the pointers to geometric elements proposed in early SDW models. In fact, a feature identifier does not denote a geometry, rather an entity that has also a semantics. Therefore some spatial operations, such as the spatial merge when applied to features, have a semantic value besides a geometric one. In the examples that will follow, spatial objects are indicated by their names.

Basic Concepts

To introduce the notion of schema and cube, we first need to define the following notions: *domain*, *level*, *level hierarchy*, *dimension* and *measure*. Consider the concept of domain. A domain defines the set of values that may be assigned to a property of facts, that is to a measure or to a dimension level. The domain may be single-valued or multi-valued; it may be spatial or non-spatial. A formal definition is given as follows.

Definition 2 (Domain and spatial domain): Let V be the set of values and F the set of features with $F \subseteq V$. A domain Do is single-valued if $Do \subseteq V$; it is multi-valued if $Do \subseteq 2^V$, in which case the elements of the domain are subsets of values. Further, the domain Do is a single-valued spatial domain if $Do \subseteq F$; it is a multi-valued spatial domain if $Do \subseteq 2^F$. We denote with DO the set of domains $DO = \{Do_1, \dots, Do_k\}$.

Example 1: In the road accident SDW, the single-valued domain of the property victims is the set of positive integers. A possible spatial domain for the position of the accidents is the set $\{a4, a5, s35\}$ consisting of features which represent roads. We stress that in this example the position is a feature and not a mere geometric element, e.g., the line representing the geometry of the road.

The next concept we introduce is that of *level*. A level denotes the single level of granularity of both dimensions and measures. A level is defined by a name and a domain. We also define the notion of partial ordering among levels, which describes the relationship among different levels of detail.

Definition 3 (Level): A level is a pair $\langle Ln, Do \rangle$ where Ln is the name of the level and Do its domain. If the domain is a spatial domain, then the level is spatial; otherwise it is non-spatial.

Let L_{v1} and L_{v2} be two levels, $\text{dom}(L_v)$ the function returning the domain of level L_v , and $\leq_v$ a partial order over V . We say that $L_{v1} \leq_v L_{v2}$ iff for each $v1 \in \text{dom}(L_{v1})$, it exists $v2 \in \text{dom}(L_{v2})$ such that $v1 \leq_v v2$. We denote with L_V the set of levels. The relationship $L_{v1} \leq_v L_{v2}$ is read: L_{v1} is less coarse (or more detailed) than L_{v2} .

Example 2: Consider the following two levels: $L_1 = \langle \text{AccidentAtLargeScale}, \text{PointAt1:1'000} \rangle$, $L_2 = \langle \text{AccidentAtSmallScale}, \text{PointAt1:50'000} \rangle$. Assume that $\text{Do}_1 = \text{PointAt1:1'000}$ and $\text{Do}_2 = \text{PointAt1:50'000}$ are domains of features representing accidents along roads at different scales. If we assume that $\text{Do}_1 \leq_v \text{Do}_2$ then it holds that $\text{AccidentAtLargeScale} \leq_v \text{AccidentAtSmallScale}$.

The notion of level is used to introduce the concept of *hierarchy of levels*, which is then applied to define dimensions and measures.

Definition 4 (Level hierarchy): Let L be a set of n levels $L = \{L_{v_1}, \dots, L_{v_n}\}$. A level hierarchy H is a lattice over L : $H = \langle L, \leq_v, L_{v_{\text{top}}}, L_{v_{\text{bot}}} \rangle$ where $\leq_v$ is a partial order over the set L of levels, and $L_{v_{\text{top}}}$, $L_{v_{\text{bot}}}$, respectively, the top and the bottom levels of the lattice.

Given a level hierarchy H , the function $\text{LevelsOf}(H)$ returns the set of levels in H . For the sake of generality, we do not make any assumption on the meaning of the partial ordering. Further, we say that a level hierarchy is of type *spatial* if all the levels in L are spatial; *non-spatial* when the levels are non-spatial; *hybrid* if L consists of both spatial and non-spatial levels. This distinction is analogous to the one defined by Han et al. (1998).

Example 3: Consider again the previous example of hierarchy of administrative entities. If the administrative entities are described by spatial features and thus have a geometry, then they form a spatial hierarchy; if they are described simply by names, then the hierarchy is non-spatial; if some levels are spatial and others are non-spatial, then the hierarchy is hybrid.

At this point we introduce the concepts of *dimensions*, *measures* and *spatial measures*. *Dimensions* and *spatial measures* are defined as hierarchies of levels. Since there is no evidence that the same concept is useful also for numeric measures, we introduce the notion of hierarchy only for the measures that are spatial. Further, as we assume that measures can be assigned subset of values, the domain of a (spatial) measure is multivalued.

Definition 5 (Dimension, measure and spatial measure): We define:

- A dimension D is a level hierarchy. The domains of the dimension levels are single-valued. Further, the hierarchy can be of type: spatial, non-spatial and hybrid;
- A measure M is defined by a unique level $\langle M, Do \rangle$, with Do a multi-valued domain; and
- A spatial measure SM is a level hierarchy. The domains of the levels are multi-valued. Moreover the level hierarchy is spatial.

To distinguish the levels, we use the terms *dimension* and *spatial measure levels*. Note that the levels of the spatial measure are all spatial since we assume that the locations of facts can be represented at granularities that have a geometric meaning. Finally, we introduce the concept of *multigranular spatial schema* to denote the whole structure of the SDW.

Definition 6 (Multigranular spatial schema): A multigranular spatial schema S (schema, in short) is the tuple $S = \langle D1, ..Dn, M1, ...Mm, SM \rangle$ where:

- D_i is a dimension, for each $i = 1, \dots, n$;
- M_j is a non-spatial measure, for each $j = 1, \dots, m$; and
- SM is a spatial measure.

We assume the spatial measure to be unique in the schema. Although in principle that could be interpreted as a limitation, we believe it is a reasonable choice since it seems adequate in most real cases.

Example 4: Consider the following schema S for the road accidents SDW:
 $S = \langle \text{date}, \text{administrativeUnit}, \text{victims}, \text{location} \rangle$ where:

- $\{\text{date}, \text{administrativeUnit}\}$ are dimensions with the following simple structure:
 - $\text{date} = \langle \{\text{year}, \text{month}\}, \leq_{\text{date}}, \text{month}, \text{year} \rangle$ with $\text{month} \leq_{\text{date}} \text{year}$
 - $\text{administrativeUnit} = \langle \{\text{municipality}, \text{region}, \text{state}\}, \leq_{\text{adm}}, \text{municipality}, \text{state} \rangle$ with $\text{municipality} \leq_{\text{adm}} \text{region} \leq_{\text{adm}} \text{state}$;

- victims is a non-spatial measure;
- location is the spatial measure. Let us call $M_1 = \text{AccidentAtLargeScale}$ and $M_2 = \text{AccidentAtSmallScale}$, two measure levels representing accidents at two different scales. Then the measure is defined as follows: $\langle \{M_1, M_2\} \leq_{\text{pos}}, M_1, M_2 \rangle$ such that $M_1 \leq_{\text{pos}} M_2$.

Finally, we introduce the concept of *cube* to denote the extension of our SDW. A cube is a set of cells containing the measure values defined with respect a given granularity of dimensions and measures. To indicate the level of granularity of dimensions, the notion of *schema level* is introduced. A schema level is a schema limited to specific levels. A cube is thus defined as an instance of a schema level.

Definition 7 (Schema level): Let $S = \langle D_1, \dots, D_n, M_1, \dots, M_m, SM \rangle$ be a schema. A schema level SL for S is a tuple: $\langle DLv_1, \dots, DLv_n, M_1, \dots, M_m, Slv \rangle$ where:

- $DLv_i \in \text{LevelsOf}(D_i)$, is a level of dimension D_i (for each $i = 1, \dots, n$);
- M_i is a non-spatial measure (for each $i = 1, \dots, m$); and
- $Slv \in \text{LevelsOf}(SM)$ is a level of the spatial measure SM

Since non-spatial measures have a unique level, they are identical in all schema levels. The cube is thus formally defined as follows:

Definition 8 (Cube and state): Let $SL = \langle DLv_1, \dots, DLv_n, M_1, \dots, M_m, Slv \rangle$ be a schema level.

A cube for SL, C_{SL} is the set of tuples (cells) of the form: $\langle d_1, \dots, d_n, m_1, \dots, m_m, sv \rangle$ where:

- d_i is a value for the dimension level DLv_i ;
- m_i is a value for the measure M_i ; and
- sv is the value for the spatial measure level Slv .

A state of a SDW is defined by the pair $\langle SL, C_{SL} \rangle$ where SL is a schema level and C_{SL} a cube.

The *basic cube* and *basic state* respectively denote the cube and the schema level at the maximum level of detail of the dimensions and spatial measure.

Example 5: Consider the schema *S* introduced in example 4 and the schema level <month, municipality, victims, accidentAtLargeScale>. An example of fact contained in a cube for such a schema level is the tuple <May 2005, Milan, 20, A4> where the former two values are dimension values and the latter two values are measure values. In particular, A4 is the feature representing the location at the measure level accidentAtLargeScale.

Spatial OLAP

After presenting the representation constructs of the model, we introduce the spatial OLAP operators. In order to motivate our choices, we first discuss three kinds of requirements that the concept of hierarchy of measures poses on these operators and thus the assumptions we have made.

Requirements and Assumptions

Interrelationship Between Dimensions and Spatial Measures

A first problem due to the introduction of the hierarchy of measures may be stated in these terms: Since a measure level is functionally dependent on dimensions, is this dependency still valid if we change the granularity of the measure? Consider the following example: assume the cube in example 4 and consider an accident that occurred in May 2005 in the municipality of Milan, located in point *P* along a given road, and having caused two victims. Now assume a decrease in the granularity of the position, thus representing the position no longer as a point but as a portion of road. The question is whether the dimension values are affected by such a change. We may observe that both cases are possible: (a) The functional dependency between a measure and a dimension is not affected by the change of spatial granularity of the measure if the dimension value does not depend on the geometry of the measure. This is the case for the dimension *date of accident*; since the date of an accident does not depend on the geometry of the accident, the dimension value does not change with the granularity. In this case we say that the date dimension is *invariant*; (b) The opposite case occurs if a spatial relationships exists between the given dimension and the spatial measure. For example, in the previous example, since it is reasonable to assume that a relationship of spatial containment is implicitly defined between the administrative unit and the accident, if the granularity of

position changes, say the position is expressed not by a point but a line, it may happen that the relationship of containment does not hold any longer. In such a case, the value of the dimension level would vary with the measure of granularity. Since this second case entails complex modelling, in order to keep the model relatively simple, we assume that all dimensions are invariant with respect to spatial measure granularity. Therefore, all levels of a spatial measure have the same functional dependency from dimensions.

Aggregation of Spatial Measures

The second issue concerns the operators for the spatial aggregation of spatial measures. Such operators compute, for example, the union and intersection of a set of geometries, the geometry with maximum linear or aerial extent out of a set of one-dimensional and two-dimensional geometries and the MBB (Minimum Bounding Box) of a set of geometries. In general, in the SDW literature these operators are supposed to be applied only to geometries and not to features. Moreover, as previously remarked, a standard set of operators for spatial aggregation has not been defined yet.

For the sake of generality, in our model we do not make any choice about the set of possible operations. We only impose, since we allow representing spatial measures as features, that the operators are applied to sets of features and return a feature. Further, the result is a new or an existing feature, depending on the nature of the operator. For example, the union (or merge) of a set of features, say states, is a newly-created feature whose geometry is obtained from the geometric union of the features' geometries. Notice also that the type of the result may be a newly-created feature type. In fact, the union of a set of states is not itself a state and therefore the definition of a new type is required to hold the resulting features.

Coarsening of Spatial Measures

The next issue is whether the result of a spatial aggregation can be represented at different levels of detail. If so, data analysis would become much more flexible, since the user would be enabled not only to aggregate spatial data but also to dynamically decrease their granularity. To address this requirement, we assume that the model includes not only operators for spatial aggregation but also operators for decreasing the spatial granularity of features. We call these operators *coarsening operators*. As previously stated, coarsening operators include operators for cartographic generalization (Camossi & Bertolotto, 2003) and fuzzyfication operators. A simple example of fuzzyfication is the operation mapping of a point of coordinates (x,y) into a close point by reducing the number

of decimal digits of the coordinates. These operators are used in our model for building the hierarchy of spatial measures.

When a measure value is expressed according to a lower granularity, the dimension values remain unchanged, since dimensions are assumed to be invariant. As a simple example, consider the position of an accident. Suppose that an aggregation operation, e.g., MBB computation, is performed over positions grouped by date. The result is some new feature, say *yearly accidents*, with its own polygonal geometry. At this point we can apply a coarsening operator and thus a new measure value is dynamically obtained, functionally dependent on the same dimension values. The process of grouping and abstraction can thus iterate.

Spatial Operators

Finally, we introduce the Spatial OLAP operators that are meant to support the navigation in MuSD. Since numerous algebras have been proposed in the literature for non-spatial DW, instead of defining a new set of operators from scratch, we have selected an existing algebra and extended it. Namely, we have chosen the algebra defined in Vassiliadis, 1998. The advantages of this algebra are twofold: It is formally defined, and it is a good representative of the class of algebras for cube-oriented models (Vassiliadis, 1998; Vassiliadis & Sellis, 1999), which are close to our model.

Besides the basic operators defined in the original algebra (LevelClimbing, Packing, FunctionApplication, Projection and Dicing), we introduce the following operators: MeasureClimbing, SpatialFunctionApplication and CubeDisplay. The *MeasureClimbing* operator is introduced to enable the scaling up of spatial measures to different granularities; the *SpatialFunctionApplication* operator performs aggregation of spatial measures; *CubeDisplay* simply visualizes a

Table 1. C_b = Basic cube

Month	Location	Victims
Jan 03	P1	4
Jeb 03	P2	3
Jan 03	P3	3
May 03	P4	1
Feb 04	P5	2
Feb 04	P6	3
Mar 04	P7	1
May 04	P8	2
May 04	P9	3
May 04	P10	1

Table 2. Cube 1

Year	Location	Victims
03	P1	4
03	P2	3
03	P3	3
03	P4	1
04	P5	2
04	P6	3
04	P7	1
04	P8	2
04	P9	3
04	P10	1

cube as a map. The application of these operators causes a transition from the current state to a new state of the SDW. Therefore the navigation results from the successive application of these operators.

Hereinafter we illustrate the operational meaning of these additional operators. For the sake of completeness, we present first the three fundamental operators of the native algebra used to perform data aggregation and rollup.

In what follows, we use the ensuing conventions: *S* indicates the schema, and *ST* denotes the set of states for *S*, of the form $\langle SL, C \rangle$ where *SL* is the schema level $\langle DLv_1, \dots, DLv_i, \dots, DLv_n, M_1, \dots, M_m, Slv \rangle$ and *C*, a cube for that schema level. Moreover, the dot notation *SL.DLv_i* is used to denote the *DLv_i* component of the schema level. The examples refer to the schema presented in Example 4 (limited to one dimension) and to the basic cube reported in Table 1.

Level Climbing

In accordance with the definition of Vassiliadis, the LevelClimbing operation replaces all values of a set of dimensions with dimension values of coarser dimension levels. In other terms, given a state $S = \langle SL, C \rangle$, the operation causes

Table 3. Cube 2

year	Location	#Victims
03	{P1,P2,P3,P4}	{4,2,3,1,2,1}
04	{P5,P6,P7,P8,P9,P10}	{3,3,1,3}

a transition to a new state $\langle SL', C' \rangle$ in which SL' is the schema level including the coarser dimension level, and C' is the cube containing the coarser values for the given level. In our model, the operation can be formally defined as follows:

Definition 9 (LevelClimbing): The LevelClimbing operator is defined by the mapping: LevelClimbing: $ST \times D \times LV \rightarrow ST$ such that, given a state SL , a dimension D_i and a level lv_i of D_i , $\text{LevelClimbing}(\langle SL, Cb \rangle, D_i, lv_i) = \langle SL', Cb \rangle$ with $lv_i = SL'.Dlvi$.

Example 6: Let SL be the following schema levels: $SL = \langle \text{Month}, \text{AccidentPoint}, \text{Victims} \rangle$. Cube 1 in Table 1 results from the execution of $\text{Level_Climbing}(\langle SL, \text{Basic_cube} \rangle, \text{Time}, \text{Year})$.

Packing

The Packing operator, as defined in the original algebra, groups into a single tuple multiple tuples having the same dimension values. Since the domain of measures is multi-valued, after the operation the values of measures are sets. The new state shows the same schema level and a different cube. Formally:

Definition 10 (Packing): The Packing operator is defined by the mapping: Packing: $ST \rightarrow ST$ such that $\text{Packing}(\langle SL, C \rangle) = \langle SL, C' \rangle$

Example 7: Cube 2 in Table 3 results from the operation: Pack ($SL, \text{Cube1}$)

FunctionApplication

The FunctionApplication operator, which belongs to the original algebra, applies an aggregation function, such as the standard avg and sum, to the non-spatial measures of the current state. The result is a new cube for the same schema level. Let M be the set of non-spatial measures and AOP the set of aggregation operators.

Table 4. Cube 3

year	#Victims	Location
03	13	Area1
04	10	Area2

Deûnition 11 (FunctionApplication): The FunctionApplication operator is defined by the mapping: FunctionApplication: $ST \times AOP \times M \rightarrow ST$, such that denoting with $op(C, Mi)$ the cube resulting from the application of the aggregation operator op to the measure Mi of cube C , FunctionApplication($\langle DLv_1, \dots, DLv_n, M_1, \dots, Mi, \dots, M_m, Slv \rangle, op, Mi$) = $\langle SL, C' \rangle$ with cube $C' = op(C, Mi)$.

SpatialFunctionApplication

This operator extends the original algebra to perform spatial aggregation of spatial measures. The operation is similar to the previous Function Application. The difference is that the operator is meant to aggregate spatial measure values.

Definition 12 (SpatialFunctionApplication): Let SOP be the set of spatial aggregation operators. TheSpatialFunctionApplication operator is defined by the mapping:

SpatialFunctionApplication: $ST \times SOP \rightarrow ST$ such that, denoting with $op(C, Slv)$ the cube resulting from the application of the spatial aggregation operator sop to the spatial measure level Slv of cube C , SpatialFunctionApplication($\langle DLv_1, \dots, DLv_n, M_1, \dots, M_m, Slv \rangle, sop$) = $\langle SL, C' \rangle$ with $C' = sop(C, Slv)$.

Example 8: Cube 3 in Table 4 results from the application of two aggregation operators, respectively on the measures victims and AccidentPoint. The result of the spatial aggregation is a set of features of a new feature type.

Measure Climbing

The MeasureClimbing operator enables the scaling of spatial measures to a coarser granularity. The effect of the operation is twofold: a) it dynamically applies a coarsening operator to the values of the current spatial measure level to obtain coarser values; and b) it causes a transition to a new state defined by a schema level with a coarser measure level.

Table 5. Cube 4

Year	#Victims	FuzzyLocation
03	13	Id
04	10	Id2

Definition 13 (MeasureClimbing): Let COP be the set of coarsening operators. The MeasureClimbing operator is defined by the mapping: MeasureClimbing : ST x COP $\rightarrow$ ST such that denoting with:

- $op(Slv)$: a coarsening operator applied to the values of a spatial measure level Slv
- $SL = \langle DLv_1, \dots, DLv_i, \dots, DLv_n, M_1, \dots, M_m, Slv \rangle$
- $SL' = \langle DLv_1, \dots, DLv_i, \dots, DLv_n, M_1, \dots, M_m, Slv' \rangle$

MeasureClimbing(SL, op)=SL' with $Slv' = op(Slv)$;

Example 9: Cube 4 in Table 5 results from the application of the MeasureClimbing operator to the previous cube. The operation applies a coarsening operator to the spatial measure and thus changes the level of the spatial measure, reducing the level of detail. In Cube 4, “FuzzyLocation” is the name of the new measure level.

DisplayCube

This operator is introduced to allow the display of the spatial features contained in the current cube in the form of a cartographic map. Let MAP be the set of maps.

Definition 14 (DisplayCube): The operator is defined by the mapping: DisplayCube: ST $\rightarrow$ MAP so that, denoting with m, a map: DisplayCube($\langle SL, C \rangle$) = m.

As a concluding remark on the proposed algebra, we would like to stress that the model is actually a general framework that needs to be instantiated with a specific set of aggregation and coarsening operators to become operationally meaningful. The definition of such set of operators is, however, a major research issue.

Future Trends

Although SMD models for spatial data with geometry address important requirements, such models are not sufficiently rich to deal with more complex

requirements posed by innovative applications. In particular, current SDW technology is not able to deal with complex objects. By complex spatial objects, we mean objects that cannot be represented in terms of simple geometries, like points and polygons. Complex spatial objects are, for example, continuous fields, objects with topology, spatio-temporal objects, etc. Specific categories of spatio-temporal objects that can be useful in several applications are diverse trajectories of moving entities. A trajectory is typically modelled as a sequence of consecutive locations in a space (Vlachos, 2002). Such locations are acquired by using tracking devices installed on vehicles and on portable equipment. Trajectories are useful to represent the location of spatial facts describing events that have a temporal and spatial evolution. For example, in logistics, trajectories could model the “location” of freight deliveries. In such a case, the delivery would represent the spatial fact, characterized by a number of properties, such as the freight and destination, and would include as a spatial attribute the trajectory performed by the vehicle to arrive at destination. By analyzing the trajectories, for example, more effective routes could be detected. Trajectories result from the connection of the tracked locations based on some interpolation function. In the simplest case, the tracked locations correspond to points in space whereas the interpolating function determines the segments connecting such points. However, in general, locations and interpolating functions may require a more complex definition (Yu et al., 2004). A major research issue is how to obtain summarized data out of a database of trajectories. The problem is complex because it requires the comparison and classification of trajectories. For that purpose, the notion of trajectory similarity is used. It means that trajectories are classified to be the same when they are sufficiently similar. Different measures of similarity have been proposed in the literature (Vlachos et al., 2002). A spatial data warehouse of trajectories could provide the unifying representation framework to integrate data mining techniques for data classification.

Conclusion

Spatial data warehousing is a relatively recent technology responding to the need of providing users with a set of operations for easily exploring large amounts of spatial data, possibly represented at different levels of semantic and geometric detail, as well as for aggregating spatial data into synthetic information most suitable for decision-making. We have discussed a novel research issue regarding the modelling of spatial measures defined at multiple levels of granularity. Since spatial data are naturally available at different granularities, it seems reasonable to extend the notion of spatial measure to take account of this requirement. The MuSD model we have defined consists of a set of represen-

tation constructs and a set of operators. The model is defined at the conceptual level in order to provide a more flexible and general representation. Next steps include the specialization of the model to account for some specific coarsening operators and the mapping of the conceptual model onto a logical data model as a basis for the development of a prototype.

References

- Bedard, Y., Gosselin, P., Rivest, S., Proulx, M., Nadeau, M., Lebel, G., & Gagnon, M. (2003). Integrating GIS components with knowledge discovery technology for environmental health decision support. *International Journal of Medical Informatics*, 70, 79-94.
- Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65-74.
- Camossi, E., Bertolotto, M., Bertino, E., & Guerrini, G. (2003). A multigranular spatiotemporal data model. In *Proceedings of the ACM GIS* (pp. 94-101).
- Clementini, E., di Felice, P., & van Oosterom, P. (1993). A small set of formal topological relationships suitable for end-user interaction. In *LNCS 692: Proceedings of the 3rd International Symposium on Advances in Spatial Databases, SSD '93* (pp. 277-295).
- Fidalgo, R. N., Times, V. C., Silva, J., & Souza, F. (2004). GeoDWFrame: A framework for guiding the design of geographical dimensional schemas. In *LNCS 3181: Proceedings of the 6th International Conference on Data Warehousing and Knowledge Discovery, DaWaK 2004* (pp. 26-37).
- Fonseca, F., Egenhofer, M., Davies, C., & Camara, G. (2002). Semantic granularity in ontology-driven geographic information systems. *Annals of Mathematics and Artificial Intelligence, Special Issue on Spatial and Temporal Granularity*, 36(1), 121-151.
- Han, J., Stefanovic, N., & Kopersky, K. (1998). Selective materialization: An efficient method for spatial data cube construction. In *Proceedings of Research and Development in Knowledge Discovery and Data Mining, Second Pacific-Asia Conference, PAKDD'98* (pp. 144-158).
- Han, J., Altman R., Kumar, V., Mannila, H., & Pregibon, D. (2002). Emerging scientific applications in data mining. *Communication of the ACM*, 45(8), 54-58.
- Jensen, C., Kligys, A., Pedersen T., & Timko, I. (2002). Multidimensional data modeling for location-based services. In *Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems* (pp. 55-61).

- Kimbal, R. (1996). The data warehouse toolkit. New York: John Wiley & Sons.
- Longley, P., Goodchild, M., Maguire, D., & Rhind, D. (2001). *Geographic information systems and science*. New York: John Wiley & Sons.
- Lopez, I., & Snodgrass, R. (2005). Spatiotemporal aggregate computation: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 17(2), 271-286.
- Malinowski, E. & Zimanyi, E. (2004). Representing spatiality in a conceptual multidimensional model. *Proceedings of ACM GIS'04*, 12-21.
- Marchant, P., Brisebois, A., Bedard, Y., & Edwards G. (2004). Implementation and evaluation of a hypercube-based method for spatiotemporal exploration and analysis. *ISPRS Journal of Photogrammetry & Remote Sensing* 59, 6-20.
- Meratnia, N., & de By, R. (2002). Aggregation and Comparison of Trajectories. *Proceedings of ACM GIS'02*, 49-54.
- OGC--OpenGIS Consortium. (2001). *OpenGIS[®] abstract specification, topic 1: Feature geometry (ISO 19107 Spatial Schema)*. Retrieved from <http://www.opengeospatial.org>
- OGC—Open Geo Spatial Consortium Inc. (2003). *OpenGIS[®] reference model*. Retrieved from <http://www.opengeospatial.org>
- Papadias, D., Kalnis, P., Zhang, J., & Tao, Y. (2001). Efficient OLAP operations in spatial data warehouses. *LNCS: 2121, Proceedings of the 7th Int. Symposium on Advances in Spatial and Temporal Databases*, 443-459.
- Pedersen, T., & Tryfona, N. (2001). Pre-aggregation in spatial data warehouses. *LNCS: 2121, Proceedings of the 7th Int. Symposium on Advances in Spatial and Temporal Databases*, 460-480.
- Rao, F., Zhang, L., Yu, X., Li, Y., & Chen, Y. (2003). Spatial hierarchy and OLAP-favored search in spatial data warehouse. *Proceedings of the 6th ACM International Workshop on Data Warehousing and OLAP, DOLAP '03*, 48-55.
- Rigaux, P., Scholl, M., & Voisard, A. (2002). *Spatial databases with applications to Gis*. New York: Academic Press.
- Rivest, S., Bedard, Y., & Marchand, P. (2001). Towards better support for spatial decision making: Defining the characteristics of spatial on-line analytical processing (SOLAP). *Geomatica*, 55(4), 539-555.
- Savary, L., Wan, T., & Zeitouni, K. (2004). Spatio-temporal data warehouse design for human activity pattern analysis. *Proceedings of the 15th International Workshop On Database and Expert Systems Applications (DEXA04)*, 81-86.

- Scotch, M., & Parmantoa, B. (2005). SOVAT: Spatial OLAP visualization and analysis tools. *Proceedings of the 38th Hawaii International Conference on System Sciences*
- Shekhar, S. , Lu. C. T., Tan, X., Chawla, S., & Vatsavai, R. (2001). Map cube: A visualization tool for spatial data warehouse. In H. J. Miller & J. Han (Eds.), *Geographic data mining and knowledge discovery*. Taylor and Francis.
- Shekhar, S., & Chawla, S. (2003). *Spatial databases: A tour*. New Jersey: Prentice Hall.
- Spaccapietra, S., Parent, C., & Vangenot, C. (2000). GIS database: From multiscale to multirepresentation. In B. Y.Choueiry and T. Walsh (Eds.), *Abstraction, reformulation, and approximation, LNAI 1864. Proceedings of the 4th International Symposium, SARA-2000*, Horseshoe Bay, Texas, USA.
- Theodoratos, D., & Sellis, T. (1999). Designing data warehouses. *IEEE Transactions on Data and Knowledge Engineering*, 31(3), 279-301.
- Vassiliadis, P. , & Sellis, T. (1999). A survey of logical models for OLAP databases. *ACM SIGMOD Record*, 28(4), 64-69.
- Vassiliadis, P. (1998). Modeling multidimensional databases, cubes and cube operations. *Proceedings of the 10th Scientific and Statistical Database Management Conference (SSDBM '98)*, 53-62.
- Vlachos, M., Kollios, G., & Gunopulos, D.(2002). Discovering similar multidimensional trajectories. *Proceedings of 18th ICDE*, 273-282.
- Wang, B., Pan, F., Ren, D., Cui, Y., Ding, D. et al. (2003). Efficient olap operations for spatial data using peano trees. *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 28-34.
- Worboys, M. (1998). Imprecision in finite resolution spatial data. *GeoInformatica*, 2(3), 257-279.
- Worboys, M., & Duckam, M. (2004). *GIS: A computing perspective* (2nd ed.). Boca Raton, FL: CRC Press.
- Yu, B., Kim, S. H., Bailey, T., & Gamboa R. (2004). Curve-based representation of moving object trajectories. *Proceedings of the International Database Engineering and Applications Symposium, IDEAS 2004*, 419-425.
- Zhang, D., & Tsotras, V. (2005). Optimizing spatial Min/Max aggregations. *The VLDB Journal* , 14, 170-181.