

Implementation and evaluation of a hypercube-based method for spatio-temporal exploration and analysis

Pierre Marchand, pierre.marchand@scg.ulaval.ca¹

Alexandre Brisebois, alexandre.brisebois@scg.ulaval.ca¹

Yvan Bédard, yvan.bedard@scg.ulaval.ca¹

Geoffrey Edwards, geoffrey.edwards@geoide.ulaval.ca¹

¹ Center for Research in Geomatics (CRG) & Geomatics for Informed Decisions (GEOIDE)

Pavillon Casault - Université Laval, Ste-Foy, G1K 7P4 Qc, Canada

Keywords : spatio-temporal topological operator dimension, spatial dimension, spatial measure, multidimensional approach, olap, Automatic Position Reporting System, cognition.

1) Abstract

In this article we test and evaluate the implementation of a dimension of spatio-temporal topological operators within a hypercube, a multidimensional database (MDDB) structure formed by the conjunction of several dimensions. Our goal is to permit adequate Spatio-Temporal Exploration and Analysis (STEA) in the context of Automatic Position Reporting System (APRS), the amateur radio system for position report transmission. Mobile APRS stations are equipped with GPS navigation systems to provide real-time positioning reports. The multidimensional approach has already demonstrated a good potential for spatio-temporal exploration and analysis despite its lack of explicit spatial, temporal and spatio-temporal topological operators that are of key importance for spatio-temporal knowledge discovery. Our approach implements spatio-temporal topological operators in multidimensional databases (MDDBs) through a hierarchy of topological operators representing spatial and temporal relationships between instances of objects, this hierarchy covers the three possible domains of spatio-temporal topological constraints *e.g.* spatial, temporal and spatio-temporal. At the root of the hierarchy users can make use of simple operators such as “same place”, “same time” or “same time, same place” in their multidimensional query. As they

drill down into the hierarchy more detailed topological operators are made available. This hierarchy is structured according to multiple levels of granularity based on cognitive models, generalized relationships and formal models of topological relationships. In this article we demonstrate that our generic approach allows adequate STEA by permitting spatio-temporal topological analysis within the multidimensional approach. We also demonstrate that such an implementation offers query run times which permit to maintain a “train-of-thought” during exploration and analysis as they are compatible with Newell’s cognitive band (query runtime <10 seconds) (Newell, 1990).

2) Introduction

Spatio-temporal exploration and spatio-temporal analysis are two complementary processes of spatio-temporal knowledge discovery, a part of Exploratory Data Analysis (EDA, as defined by Glymour et al., 1997) dealing with spatio-temporal data (Marchand et al., 2002). The *exploration process* identifies hypotheses that are of interest to the user. Hypotheses usually consist of potential patterns, associations and unusual occurrences which help determine prediction rules, clusters, etc. The *analysis process* determines the validity of the hypotheses formulated in the exploration process and, in some cases, proposes new ones. The EDA iterative navigation from general views (coarse granularity) to very detailed views (fine granularity) according to multiple themes implies running large queries against large volumes of data in reasonable time. For spatio-temporal databases, aside from performance issues, these procedures depend on SQL which has always been too abstruse for typical end users due in part to numerous spatial and temporal extensions *cf.* (Cardenas and Thomas, 1988; Egenhofer and Herring, 1993; Merret et al., 2002). SQL offers poor support for basic analysis operations such as comparisons which are difficult to specify (Kimball and Strehio, 1997). Users are also required to understand and master the logic structure of the spatio-temporal database. This structure still remains complex for typical end users since (Langran, 1992)’s conclusions.

The multidimensional approach has demonstrated a strong potential for efficient analysis and exploration ever since the analytical strength of matrix algebra gave birth to multidimensional analytical software tools. This potential has been confirmed by (Caron, 1998; Rivest, 2000) in the context of STEA and by (Bédard et al., 2000; Han, 1998) in the context of spatio-temporal knowledge discovery. The multidimensional approach unifies a category of principles and technologies oriented towards supporting sophisticated multidimensional analyses. The ability to simultaneously aggregate across many sets of dimensions lies at the core of the multidimensional approach. In order to efficiently achieve such tasks, data structures are highly denormalized and aggregated along conceptual hierarchies called dimensions. The potential of the multidimensional approach has been largely implemented, used, validated and made profitable (Berson and J., 1997; Date, 2000; Thomsen, 1997) since its proposition by Codd in 1993. Despite compelling advantages, this approach has not been exhaustively evaluated within the geomatics sciences (Bédard et al., 1997; Caron, 1998; Gonzales, 1999; Mattos and Zeidenstein, 1999). While it is receiving increasing interest from the geospatial knowledge discovery community, it is only recently that the GIS community has become involved.

(Bédard et al., 2000) have introduced the role to be played by the multidimensional approach within spatial data warehousing as well as geospatial principles for multidimensional databases (MDDBs). (Rivest et al., 2001) have highlighted the need for a client interface which permits representation and manipulation of geometric attributes. (Bédard et al., 1997) coined the term SOLAP (Spatial OLAP) to define a new category of tools which they defined as *“a visual platform built especially to support rapid and easy spatio-temporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays”*. (Clementini et al., 2000; Stefanovic, 1997; Stefanovic et al., 2000) proposed methods for computation of spatial data cubes such as approximation and selective materialization. (Rivest et al., 2001) and (Rivest, 2000) defined the conceptual foundations, associated concepts, essential features, desirable

characteristics and presented application prototypes. (Marchand et al., 2002) described the theoretical justification for applying the multidimensional approach to the geomatics sciences and introduced the Spatio-Temporal Topological Operator Dimension (Spatio-Temporal TOD). This dimension implements a hierarchy of spatio-temporal topological operators in a hypercube within a MDDDB to permit adequate spatio-temporal exploration and analysis.

In its simplest form, the multidimensional approach is structured according to a three-tier client-server architecture. A multidimensional database is hosted by a server which is accessed by a client. Multidimensional databases (MDDDBs) are made of cubes, dimensions, hierarchies and measures. A cube is a multidimensional array resulting from the conjunction of several dimensions. For example, in a sales application, typical dimensions are products, time, locations and customers (Marchand et al., 2002). The cube's dimensions are usually organized into conceptual hierarchies, which represent distinct categories, and grouped into levels. Members of a level are usually aggregated to form their parent's value according to aggregation rules *e.g.* in a temporal dimension, days roll into months, and months roll into quarters. Measures are the quantitative values to be analyzed in the MDDDB. For example, for a sales application, typical measures are revenue, cost, units sold, discounts and returns (Marchand et al., 2002). Multidimensional views are produced when measures are analyzed against the different dimension categories of a cube. A typical multidimensional query for a sales application would be : how much revenue (the measure) did a company have last month (a member of the Time dimension) in Europe (a member from the Locations dimension) for all of its products (the top level of the Product dimension) and all of its customers (the top level of the Customer dimension). Usually, any multidimensional view (multidimensional query) can either be built or retrieved by an MDDDB client in a few seconds and displayed through multiple types of graphical representations (Marchand et al., 2002; Thomsen, 1997). Figure 1 presents a visual representation of a multidimensional query applied to a hypercube.

3) The Spatio-Temporal Topological Operator Dimension (Spatio-temporal TOD)

The Spatio-Temporal Topological Operator Dimension (Spatio-temporal TOD) implements spatio-temporal topological operators in hypercube within MDDBs to permit adequate spatio-temporal exploration and analysis (Marchand et al., 2002). It involves a hierarchy of topological operators based on spatial and temporal relationships between instances of spatial objects and temporal primitives. These operators cover the three possible domains of spatio-temporal topological constraints *e.g.* spatial, temporal and spatio-temporal. At the root of the hierarchy users can make use of simple operators such as “same place”, “same time” or “same time, same place” in their multidimensional query. Figure 2 presents an example of a business query implying the Spatio-Temporal TOD to isolate the stores locations (polygons) adjacent in space and over-lapped-by in time to trucking delivery routes (lines) in Belgium for milk products during the third quarter of 2001. This figure also illustrates the successive refinements that can be made through the Spatio-Temporal TOD to achieve such spatial and temporal topological constraints between stores and routes.

Within the Spatio-Temporal TOD, the spatial topological relationships are described through the DE+9IM proposed by (Clementini and Di Felice, 1994) and by the 9-intersection method (9IM) proposed by (Egenhofer and Herring, 1994) as well as generalizations used by (Normand, 1999) or defined by the present author. The temporal topological relationships are described through the DE+9IM proposed by (Clementini and Di Felice, 1994), by the 9-intersection method (9IM) proposed by (Egenhofer and Herring, 1994), by (Allen, 1983)’s temporal interval algebra as well as generalizations defined by the present author. Stored in a MDDB hypercube as 9-digit strings (which are equivalent to the input string of either the eRelate or cRelate ISO TC211 topological operators) these relationships make up the leafs of the Spatio-Temporal TOD hierarchy. The hierarchy of the Spatio-Temporal TOD is structured according to

multiple levels of topological relationships granularities and to behavior models of spatial analysis to improve database usability. This approach provides faster and more intuitive spatio-temporal analysis in MDDBs. It can be inserted in any multidimensional cube containing spatial and/or temporal data and is software independent. Preliminary results from (Marchand et al., 2002) have shown that this approach could well be implemented in a reasonable timeframe.

4) Principles of spatio-temporal multidimensional exploration and analysis

4.1. Introduction

A multidimensional query is formed by the conjunction of one or more dimensions members with a measure. Applied to one or more hypercubes of a MDDB, this set of constrains results in a cross-dimensional view that can be represented by tables, histograms, scatter plots, cartographic maps, *etc.* Exploration of the multidimensional database is carried out by navigating along the dimension hierarchies and browsing through the measures. (Wang et al., 2000) proposed eight rules which promote the ability to use fast and easy to understand graphic displays (such as maps, tabular displays, pie charts, bar charts, line graphs, histograms, *etc*) within an environment. The multidimensional approach complies with this set of requirements. Two concepts have been proposed by (Stefanovic, 1997) and later by (Bédard et al., 2000) to enrich the multidimensional approach : spatial dimension and spatial measure. These concepts aim at integrating a spatial perspective within the multidimensional approach in order to improve analysis, design and usage of spatial multidimensional databases. Prior to supporting STEA with the multidimensional approach one must understand approaches permitting the identification of change as well as their underlying requirements. Two main approaches guide the identification of change in spatio-temporal data ; inductive and deductive methods.

4.2. Identification of spatio-temporal change

Two main approaches guide the identification of change in spatio-temporal data. When the focus is on the evolution of individual entities or small sets of entities, deductive methods are used. These methods are based on mathematical and qualitative representations of geographical changes *i.e. a priori* knowledge of possible spatio-temporal relationships. As the definition of models and proofs currently involves mathematical complexity, qualitative reasoning methods are limited to the representation of a small number of geographical entities (Therault et al., 1999). When the focus is on the evolution of large sets of entities, inductive methods are used. These methods are based on an observation and statistical analysis of a geographical phenomenon (*e.g.*, point pattern analysis, spatial autocorrelation, centrographic analysis). Both these methods span over a large set of disciplines ranging from environmental to urban and biological studies (Therault et al., 1999). Usage of the multidimensional approach in the context of STEA is independent of these approaches as users browse a multidimensional database through a client supporting multiple modes of graphic representations (Rivest et al., 2001) built according to custom or predefined semiology rules (Thomsen, 1997). As with the relational approach, it is supposed that the MDDB has been designed according to the principles of the domain of interest. These factors insure that users have a complete control over design and usability of multidimensional databases according to the principles of their approach. Nevertheless, as the Spatio-Temporal TOD results from the combination of constraints subsets from an initial general set we can infer that the resulting spatio-temporal reasoning follows a qualitative, constraint-driven, approach (see (El-Geresy et al., 2000)). Spatio-temporal reasoning is achieved here through the combination of the spatial and temporal topological relationships under the Spatio-Temporal TOD constraint hierarchy.

4.3. Spatial dimensions

(Stefanovic, 1997) proposed that dimensions which hold geometric spatial reference in their hierarchies be called spatial dimensions. (Stefanovic, 1997) and (Bédard et al., 2000) consider three types of spatial dimensions based on the content and nature of the geometric spatial reference members of the dimension : 1) non-geometric, 2) geometric-to-non-geometric and 3) fully geometric.

A non-geometric spatial dimension is a hierarchy containing only non-geometric nominal members which are used to position a phenomena in space. Such dimensions are frequently encountered in business applications which make use of geographical breakdowns in combination with other categories such as time or product lines (Gonzales, 1999). (Caron, 1998) has identified the capacities and limitations of such dimensions for spatio-temporal exploration. Non-geometric spatial dimensions can be implemented in any multidimensional architecture as long as cartographic representations and navigation are not required (Bédard et al., 2000). It is important to note that the absence of cartographic representations and navigation does not refrain users from carrying out spatial cognition during exploration and analysis as nominal references of spatial entities participate in spatial cognition through language. A geometric-to-non-geometric spatial dimension is a hierarchy whose root/leaf level members have a geometric representation but whose generalization become non-geometric at a certain level of aggregation (Rivest et al., 2001). The benefit of this approach resides in the simplification potential of the measurement scales (Chevalier and Bédard, 1991) which illustrate, for example, how qualitative measurements carry less details than quantitative (Bédard et al., 2000). Geometric-to-non-geometric spatial dimensions can only be implemented in multidimensional architectures as long as cartographic representations and navigation is supported by a multidimensional database client. A fully geometric spatial dimension is a hierarchy whose members all have a geometric representation. Fully geometric spatial dimensions can be implemented in multidimensional architectures as long as cartographic representations and navigation is supported is by a multidimensional database client.

There is evidence that spatial cognition relies on hierarchies to support representations of space (Golledge, 1992; Remolina et al., 1999; Taylor and Tversky, 1992). Nevertheless, as the relevance of the spatial dimension concept has not yet been demonstrated it should be employed with caution. It can be used to support spatial multidimensional database design, to simplify document design, to facilitate programming and maintenance. The spatial dimension concept is essential to the advent of the multidimensional approach in the geomatics community. As there is still no agreement on formalism for modeling multidimensional databases it is difficult to propose a way to identify and model spatial dimensions. We propose that the SPVL approach (Spatial Plug-in for Visual Languages (Bédard, 1999)) be used for this task. Based on ISO TC211 standards, this approach uses a font which includes pictograms used to model basic spatial constructs as well as their spatial and temporal variations. This approach is simple to use and has a high expressive power. It can be used in any formalism, as long as consistency is maintained with the model and formalism. Pictograms correspond to object classes or attributes in relational database design. For the multidimensional database design, pictograms should be associated with hierarchy levels of spatial dimensions in order to determine desired cartographic representations for each relevant level of the spatial dimension hierarchy. Despite the lack of an extended dictionary supporting geospatial metadata standards as in Perceptory (Bédard, 1999) we believe that this solution offers substantial grounds for modeling multidimensional databases which include spatial dimensions.

The spatial representation of members from each hierarchy level can result from spatial aggregations or generalizations of other level members or be completely independent. In most cases, hierarchy levels and members do not comply with topological or spatial constraints. Therefore, spatial analysis through drill downs or drill-ups of spatial dimension hierarchy levels is not possible and is misleading if conducted. Conventional multidimensional database aggregative functions aggregate measures along hierarchies. They cannot yet perform spatial aggregations or similar spatial operations during hypercube construction.

4.4. *Spatial measure*

Measures are the numeric values available for analysis in a multidimensional database. They can be perceived as quantifiable dimensions. They are sometimes called variables, a synonym for measure. They represent the numeric data of primary interest to users.

(Stefanovic, 1997) proposed that measures of multidimensional databases which include spatial representations be materialized and called spatial measures. (Stefanovic, 1997) defined a spatial measure as “*a measure which contains a collection of pointers to spatial objects*” (Stefanovic, 1997). In his experimentation (Stefanovic, 1997) built a hypercube with three dimensions (*temperature, precipitation, time*), a spatial dimension (*region name*) and three measures (*region_map, area, count*). Of the three measures, *region_map* is the only spatial measure as it contains a collection of spatial pointers pointing to the corresponding regions (Stefanovic, 1997). The measure *area* is a numerical measure representing the sum of the total areas of the corresponding spatial objects. The measure *count* is a numerical measure representing the total number of base regions. Such a configuration serves two purposes ; to provide measures (*area* and *count*) to users for exploration and analysis and to provide a link (*region_map*) between portions of the hypercube and their spatial representations. According to (Stefanovic, 1997) this link is required to dynamically perform roll-ups and drill-downs operations as it identifies spatial objects that have to be dynamically merged for representation purposes.

While we agree with these requirements in the context of a multidimensional database which builds its cartographic representation dynamically we feel it is not appropriate to associate the spatial measure concept with such a restrictive context. Spatial data warehouses often share a common spatial data layer between transactional and multidimensional databases. According to the architecture of the spatial data warehouse, this spatial data layer can be either pre-defined, built dynamically or result of a combination of both to serve specific tasks such as representation of spatial information in cartographic displays. Despite

huge achievements in the field of generalization over the last decades, several researchers are proposing multiple representation as the alternative to generalization for spatial data warehouses.

The context of (Stefanovic, 1997)'s experimentation is too restrictive to propose and demonstrate the relevance of the spatial measure concept. Nevertheless, we believe it is relevant for spatial multidimensional databases to distinguish a measure associated with spatial representations from a regular measure. This assertion is justified by the supplementary efforts required to provide adequate representation in cartographic displays as well as by the needs to provide support for adequate spatial analysis capacities through data structures such as Spatio-Temporal TOD (Marchand et al., 2002). These requirements aim at harnessing the spatial knowledge potentially embedded in spatial data structures. This spatial knowledge can be of critical importance for business (Gonzales, 1999; Mattos and Zeidenstein, 1999). Such issues have direct impact on design, implementation and maintenance of spatial multidimensional databases. It is therefore essential to adequately define this concept.

We define a spatial measure as a measure which is mapped to at least one spatial dimension whose members hold spatial representation (geometric-to-non-geometric and fully geometric spatial dimensions). We distinguish two types of spatial measures depending on the type of spatial dimension they map; geometric-to-non-geometric spatial measure and fully geometric spatial measure. A geometric-to-non-geometric spatial measure is a spatial measure which is mapped by at least one geometric-to-non-geometric spatial dimension. A fully geometric spatial measure is a spatial measure which is only mapped by fully geometric spatial dimensions. There is no non-geometric spatial measure as non-geometric spatial dimension holds only non-geometric nominal members.

Not all measures are mapped to every member of every level of every dimension. This would result in database explosion (Pendse, 2001). In most cases, few measures are mapped to all dimensions members. Sparsity quantifies the amount of sparse data in multidimensional databases. In most cases, there is a high

level of sparsity in multidimensional databases. (Pendse, 2001) presents a good introduction to the problems of database explosion and sparsity.

At present time, conventional multidimensional database aggregative functions aggregate numerical measures along hierarchies. It is likely that such functions will also perform spatial functions such as polygon amalgamation and topological analysis in the near future. In the context of spatial data mining, they could also compute clusters and even clusters in the presence of obstacles (Tung et al., 2001). Semiology could be exploited to cope with the major challenge of aggregating a large number of regions flexibly and dynamically. For example, rather than aggregating polygons to display a common spatial measure after a drill-up operation a single color could be assigned to all polygons.

Among the commercial products available at present time, ProClarity from Knosys is the only client which offers through its MapX module a true cartographic display within a multidimensional database client. This feature is made possible by linking each level member of the spatial dimension to its spatial representation through a common identification attribute. If members with spatial representations are distributed on multiple levels, then each level is associated with different spatial data coverages. This arrangement supports both geometric-to-non-geometric spatial dimensions and fully geometric spatial dimensions.

5) Experimentation

5.1. Introduction

The goal of this experimentation is to determine whether or not the Spatio-Temporal TOD permits spatio-temporal reasoning which is treated here as a qualitative, constraint-driven, approach. Our approach first defines a general set of constraints and later combines different subsets to define the different spatio-temporal relations and permit spatio-temporal reasoning through the combination of the spatial and temporal topological relationships under the Spatio-Temporal TOD constraint hierarchy. Once

implemented through the multidimensional approach, query run times should be inferior to 10 seconds in order to be compatible with (Newell, 1990)'s cognitive band. The dataset used in this experimentation is a subset of the data collected from the APRServe Network. The APRServe Network broadcasts position reports, weather reports and messages between radio amateur stations through the Internet. Stations are generally equipped with GPS navigation systems to provide real-time positioning. The dataset used in this experimentation covers the cities of New York and Washington from August 13 2001 to November 12, 2001. This dataset is made of 29.475 line objects corresponding to the linear interpolation between successive positions of 214 radio amateur mobile stations broadcasted through the Automatic Position Reporting System (ARPS). Throughout our experimentation we used a 1.7 GHz Pentium 4 computer equipped with 1Gb RAM, 160 Gb hard disk space, Oracle 9i, MS SQL Server 2000 and MS Analysis Services 2000.

5.2. Data structuration

In this experimentation we segmented the routes of every APRS mobile radio station into their smallest primitives. Each spatial segment corresponds to the linear interpolation between two successive position reports of a radio station whatever the reporting frequency. Each spatial segment is attributed a starting and finishing timestamp as well as a unique identification code. Figure 3 and table 1 present an example of this segmentation. This structure permits to topologically query, at the finest granularity, each individual spatial entity as well as its temporal primitive. These segments can later be aggregated for mobile stations or time periods once all relationships are determined.

In order to offer operators at all levels of the Spatio-Temporal TOD it is required to *a priori* determine topological relationships between spatial entities and between temporal primitives. These relationships will be aggregated upon the construction of the hypercube into generalized members according to the hierarchical relationships defined by the Spatio-Temporal TOD table. As the implementation of data into a MDDB varies according to software packages we build two tables that all MDDBs can import either as a

flat text files or as linked tables through a conventional ODBC connection. The first table contains the Spatio-Temporal TOD members as well as their hierarchical relationships expressed through parent-child relationships. The second table contains the topological relationships between spatial entities and between temporal primitives. The second table records results from the topological analysis of topological relationships between all spatial entities and between all temporal primitives. Figure 4 presents an extract from each table. These two tables are then integrated with other descriptive tables to form a stand-alone hypercube or a shared hypercube that can be used across the MDDB by to enrich multidimensional queries. There is no straightforward approach to compute all topological relationships between a set of spatial entities. In order to avoid the N^2 space requirement associated such a computation we followed a two-phase query model.

In the first phase, a spatial index is created using a spatial database such as PostGIS for PostgreSQL, Oracle Spatial, ESRI SDE or spatial dataview for Informix. This spatial index guarantees that entities which are near in the coordinate space are near in the ordering. Minimum bounding rectangles (MBRs) are then tested for intersection using the spatial index to remove the obvious cases. MBRs are coarse approximations. Testing their intersections does not guarantee intersection of actual spatial objects but highly reduces the amount of exact computations to be carried out for most common spatial datasets where most features do not overlap.

In the second phase, exact topological relationships are computed between the entities which are likely to interact using the Java Topology Suite (JTS). The Java Topology Suite is a free Java API (www.vividsolutions.com) of spatial predicates and functions based on the ISO/TC211 standards and simple features specification for SQL from the Open GIS Consortium (Marchand et al., 2002). We used the DE+9IM topological model from (Clementini and Di Felice, 1994) for testing and implementation as it corresponds to the finest granularity of the Spatio-Temporal TOD. To roll-up the topological relationships along the hierarchy, the DE+9IM was generalized into the 9IM (Egenhofer and Herring, 1994) and into

other aggregations presented in (Marchand et al., 2002). In case of a non-disjoint result, the identifier of each spatial entity from the pair is stored in the second table along with the spatial and temporal topological operator. Once all exact spatial topological relationships have been identified, Allen's temporal interval algebra (Allen, 1983) can be used to compute topological relationships between the temporal entities associated with the pair of spatial entities. Finally, the spatio-temporal topological operator is formed by the concatenation of the temporal and spatial topological operators. This spatio-temporal topological operator is joined through an intermediate table (n:m) to its appropriate level members in the first table. As in a graph, almost every member in the Spatio-Temporal TOD can have multiple parents. For every non-leaf member there are multiple ways to drill-down; spatial drill-down, temporal drill-down and spatio-temporal drill-down. Therefore, very few members are unique either on a level or throughout the hierarchy. This intermediate table permits to link facts to multiple members in the hierarchy. In order to make the Spatio-Temporal TOD available with other dimensions these two tables should be linked, *a priori* or *a posteriori*, to other descriptive tables and imported into a MDDB. Using the Spatio-Temporal TOD in conjunction with other dimensions within a MDDB permits to explore and analyze spatio-temporal data with spatio-temporal topological constraints (either cognitively compatible or based on the ISO/TC211 standards).

We recommend that the topological operators used in the second phase be of the finest granularity possible with respect to the Spatio-Temporal TOD in order to facilitate later aggregations. In order to reduce table size, we also recommend not to store spatial disjoint relationships. If there exists no relationship between two spatial entities in the second table it is assumed they are disjoint.

5.3. Implementation

In the first phase of our experimentation we used Oracle Spatial 9i to build a spatial index and identify the pairs of spatial entities which were likely to interact. It took 6 seconds to compute a spatial index (R-tree) for the 29.475 spatial entities. It took 11 minutes to identify 14.834.713 pairs of spatial entities which were

likely to interact through the SDO_FILTER. The complexity of SDO_FILTER is believed to be around $\log n$. We carried out a similar test for a larger dataset made of 853.764 spatial entities. It took 11 minutes to build the spatial index and 51 hours to identify the 1.422.288.743 pairs of spatial entities which were likely to interact. A homemade software carried out the same calculation in 38 minutes on the same computer using a fixed tile index and determinant calculation.

In the second phase we computed all relevant topological relationships between the 14.834.713 spatial pairs in less than 10 hours using the Java Topology Suite. Using Allen's temporal interval algebra (Allen, 1983), we computed all temporal topological relationships between the temporal primitives associated with the spatial entities in less than 30 minutes.

Finally, we integrated the two tables with a temporal dimension (year / month / day) and two identifying dimensions for mobile APRS stations into a hypercube in 1 hour and 10 minutes. The source relational fact table size was 6066 Mb with indexes but no aggregates. As the fact table was linked to the Spatio-Temporal TOD through an intermediate table, the original 2.438.233 facts were duplicated into 100.024.436 facts. This intermediate table permitted adequate mapping of every fact to every adequate member in the Spatio-Temporal TOD (2.385.090 members). The resulting hypercube size was 663 Mb including all possible aggregates. All multidimensional queries runtimes we monitored were inferior to 2 seconds. Figure 5 presents an example of interface which can be used to explore and analyze this spatio-temporal dataset. This interface corresponds to a MDDB client, ProClarity from Knosys, with a mapping plug-in, MapX for ProClarity.

6) Results and discussion

We successfully implemented the Spatio-Temporal TOD using conventional software and hardware. By using the Spatio-Temporal TOD alone or in conjunction with other dimensions users can impose spatio-temporal topological constraints on the multidimensional database they are exploring. The

multidimensional query runtimes are well below the upper limit of Newell's cognitive band (Newell, 1990).

These results indicate that the implementation of the Spatio-Temporal TOD in a MDDDB is both feasible and replicable. The multidimensional query runtimes are compatible with Newell's cognitive band (Newell, 1990). This guarantees that the implementation of the Spatio-Temporal TOD in a MDDDB does not hinder the users' train-of-thought. Although the Spatio-Temporal TOD is a fairly large dimension (2.385.090 members on 11 levels) it does not interfere with the functioning of the MDDDB. This can be explained by the natural ability of the multidimensional approach to deal with high levels of redundancy and sparsity. The required time for implementation is acceptable in the context of data warehousing. Frequent updates to the multidimensional database are possible and updates do not represent any challenge not addressed by data warehousing. In terms of scalability, it has been demonstrated that the multidimensional approach can support large volumes of data.

These results are consistent with those of (Soni and Kurtz, 2000). They tested the compression of a six-dimensional banking hypercube based on a 13 million-row fact table using Microsoft Analysis Services 2000. The source relational fact table size was 5188 Mb with indexes but no aggregates. The resulting hypercube size was 336 Mb even when including a significant number of aggregates. Facts inserted in a MDDDB will often be compressed by a factor varying from 20-to-1 to 10-to-1. This compression feature is built-in into most commercial MDDDB servers and its structure is proprietary. In a white paper, Microsoft presented results of an experiment where 1.2 Tb of data (700,000 products and 7.7 billion fact rows) was inserted in a single cube and offered query run times inferior to a second even in the case of complex queries (Microsoft, 2001). Processor and disk intensive operations carried out during the first and second phases are the only disadvantages of the integration of the Spatio-Temporal TOD within a MDDDB. They are nevertheless typical of the data feeding process of multilevel data warehouses and pose no specific issue not addressed by data warehousing.

There are other approaches to address the N^2 space requirement associated with the computation of topological relationships between a set of spatial entities. The plane sweep algorithm (de Berg et al., 2000) is a computational geometry algorithm which could be used in the first phase of the query model. Other solutions can be found by exploiting the explicit topological data structures of some GISs or taking advantage of topologically enriched data structures such as the Voronoi diagram (Gold et al., 1997). Finally, efficient polygon amalgamation methods (Zhou et al., 1999) and similar techniques could certainly improve efficiency. It is beyond the scope of this article to determine the adequacy of different approaches for datasets varying in size and complexity.

Clementini *et al.* proposed a hierarchy of spatial topological relations for mining spatial association rules for objects with uncertain spatial information (Clementini et al., 2000). This hierarchy is used to optimize the mining process through a progressive refinement approach. The Spatio-Temporal TOD can also be used for the same purpose in the context of spatio-temporal knowledge discovery.

7) Conclusion

In this experimentation we demonstrated that the implementation of the Spatio-Temporal TOD in a MDDB is both feasible and replicable. As expected, multidimensional query run times associated with the Spatio-Temporal TOD are contained in Newell's cognitive band (Newell, 1990) which ensures no train-of-thought loss for users.

Some spatial topological operators are available in GIS software for spatial analysis in spatio-temporal data exploration and analysis. They are processor intensive and usually require training. Through the implementation of the Spatio-Temporal TOD in a MDDB a complete, simple to use, flexible, hierarchy of spatio-temporal topological operators is available at first hand with no software training and no query processing required.

In the context of data warehousing, the results of this experimentation indicate it is possible to share the Spatio-Temporal TOD across a forest of cubes within a MDDB. The spatio-temporal topological relationships can therefore be determined only once and used from multiple cubes in conjunction with other descriptive dimensions. This perspective is very interesting for data warehouses which highly rely on geographical features *e.g.* retail, transport and environment.

The approach followed by the Spatio-Temporal TOD in this experimentation can be transposed to other types of spatial operators commonly found in GIS software. Operators such as visibility and slope can also be embedded in MDDB as dimensions. We believe it would be possible to reproduce a large share of the functionalities offered by conventional GIS software in MDDB. While the conventional mathematical and logical operators are available from the MDDB server other operators can be embedded as dimensions. The obvious advantages of such an approach is to provide spatio-temporal exploration and analysis applications that permit to maintain a train-of-thought without restrictions on the data manipulations and dataset size. The associated processing cost can be easily absorbed in the context of data warehousing.

8) Acknowledgments

We are grateful to the financial support of the GEOIDE project called GEODEM (Geospatial Decision Making) and to the Canadian Natural Sciences and Engineering Research Council. We would like to thank Eliseo Clementini from University of L'Aquila, Italy, for helping building adequate relations between levels in the Spatio-Temporal hierarchy. We thank Steve Dimse of www.findu.com for providing the experimentation dataset. We thank Nicole S. Alexander from the Spatial Products Development Group, Oracle Corporation, for her permanent support. We thank Suzie Larrivée and Clément Nolette, University of Laval, for their significant help during the experimentation. We thank Tim Merret from McGill University, Canada, for his very valuable guidelines. We also thank Martin Davis from Vivid Solutions and George Spofford from DSSLAB for their invaluable advises.

9) Bibliography

- Allen, J.F., 1983. Maintaining knowledge about temporal intervals. 26: 832--843.
- Bédard, Y., 1999. Visual modelling of spatial databases: towards spatial PVL and UML, *Geomatica*, pp. 169 to 186.
- Bédard, Y., Larrivée, S., Proulx, M.-J., Caron, P.-Y. and Létourneau, F., 1997. Geospatial Data Warehousing : positionnement technologique et stratégique.
- Bédard, Y., Merret, T. and Han, J., 2000. Fundamentals of spatial data warehousing for geographic knowledge discovery. In: H. Miller and J. Han (Editors), *Geographic data mining and knowledge discovery*.
- Berson, A. and J., S.S., 1997. *Data Warehousing, Data Mining, and OLAP*. McGraw-Hill, 612 pp.
- Cardenas, A.F. and Thomas, J., 1988. PICQUERY: a high level query language for pictorial database management. *IEEE Transactions on software engineering*, 14(5): 630-637.
- Caron, P.-Y., 1998. Étude du potentiel OLAP pour supporter l'analyse spatio-temporelle, Université Laval, Ste Foy.
- Chevalier, J.J. and Bédard, Y., 1991. Echelles de mesure spatiale et processus de modélisation appliqués au SIRS, La conférence canadienne sur les SIG, Ottawa, Canada, pp. 547-557.
- Clementini, E. and Di Felice, P., 1994. A comparison of methods for topological relationships. *Information Sciences* 80: 1-34.
- Clementini, E., Di Felice, P. and Koperski, K., 2000. Mining multiple-level spatial association rules for objects with a broad boundary. *Data & Knowledge Engineering*, 34: 251-270.
- Date, C.J., 2000. *An introduction to database systems*. Addison Wesley.
- de Berg, M., van Kreveld, M., Overmars, M. and Schwartzkopf, O., 2000. *Computational geometry : algorithms and applications*. Springer-Verlag, Berlin, 375 pp.
- Egenhofer, M.J. and Herring, J., 1994. Categorizing binary topological relations between regions, lines, and points in geographic databases. NCGIA Technical Report(94-1).
- Egenhofer, M.J. and Herring, J.R., 1993. Querying a geographical information system. In: D. Medyckyj-Scott and H.M. Hearnshaw (Editors), *Human factors in geographical information system*. Belhaven Press, London, UK, pp. 124-135.
- El-Geresy, B.A., Abdelmoty, A.I. and Jones, C.B., 2000. Episodes in space: qualitative representation and reasoning over spatio-temporal objects. *International Journal on Artificial Intelligence Tools*, 9(1): 131-152.
- Gold, C.M., Remmele, P.R. and Roos, T. (Editors), 1997. *Voronoi methods in GIS. Algorithmic Foundations of GIS. Lecture Notes in Computer Science*, 1340. Springer-Verlag, 21-35 pp.
- Golledge, R.G., 1992. Do people understand spatial concepts: the case of first-order primitives. In: A.U. Frank, Campari, I. & Formentini, U. (Editor), *Theories and methods of spatio-temporal reasoning in geographic space*. Springer-Verlag, Berlin Heidelberg, pp. 1-21.
- Gonzales, M.L., 1999. *Spatial olap: conquering geography, DB2*.

- Han, J., 1998. Olap mining: an Integration of olap with data mining. In: S. Spaccapietra and F. Maryanski (Editors), *Data Mining and Reverse Engineering: Searching for Semantics*. Chapman Hall, pp. 3-20.
- Kimball, R. and Strehio, K., 1997. Why decision support fails and how to fix it, *SIGMOD*, pp. 92-97.
- Langran, G., 1992. *Time in geographic information systems*. Taylor & Francis, London.
- Marchand, P., Bédard, Y. and Edwards, G., 2002. A hypercube-based method for spatio-temporal exploration and analysis. *GeoInformatica*(to be accepted - submitted May 2001).
- Mattos, N.M. and Zeidenstein, K., 1999. *Integrating Spatial Data with Business Data, DB2*.
- Merret, T. et al., 2002. A tutorial on database technology for geospatial applications.
- Microsoft, 2001. *T³ Project Technical Overview*, Microsoft Corporation, Redmond.
- Newell, A., 1990. *Unified theories of cognition*, 549 pp.
- Normand, P., 1999. *Modélisation des contraintes d'intégrité: théorie et exemples d'application*, Université Laval, Ste Foy, 93 pp.
- Pendse, N., 2001. *Database explosion. The OLAP report*.
- Remolina , E., Fernandez, J.A., Kuipers, B. and Gonzalez, J., 1999. Formalizing regions in the spatial semantic hierarchy: an AH-graphs implementation approach, *COSIT99*.
- Rivest, S., 2000. *Investigation des modes d'intégration physique entre un serveur de base de données multidimensionnelle et un SIG*, Université Laval, Sainte-Foy.
- Rivest, S., Bédard, Y. and 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.
- Soni, S. and Kurtz, W., 2000. *Optimizing Cube Performance Using Microsoft Analysis Services 2000*, UNISYS.
- Stefanovic, N., 1997. *Design and Implementation of On-Line Analytical Processing (OLAP) of Spatial Data*, Simon Fraser University, Vancouver, 118 pp.
- Stefanovic, N., Han, J. and Koperski, K., 2000. Object-based selective materialization for efficient Implementation of spatial data cubes. *IEEE Transactions on Knowledge and Data Engineering*, 12(6).
- Taylor, H. and Tversky, B., 1992. Descriptions and depictions of environments. *Memory and Cognition*, 20(483-496).
- Theriault, M., Claramunt, C. and Villeneuve, P.Y., 1999. A spatio-temporal taxonomy for the representation of spatial set behaviours, *Workshop on Spatio-Temporal Database Management, STDBM'99*, Edinburgh, Scotland.
- Thomsen, E., 1997. *OLAP solutions : building multidimensional information systems*. Wiley Computer Pub., New York.
- Tung, A.K.H., Hou, J. and Han, J., 2001. *Spatial Clustering in the Presence of Obstacles*, Int. Conf. on Data Engineering (ICDE'01), Heidelberg, Germany.

- Wang, B.M.Q., Kuchinsky, A. and Woodruff, A., 2000. Guidelines for using multiple views in information visualization, AVI 2000, Palermo, Italy.
- Zhou, X., Truffet, D. and Han, J., 1999. Efficient polygon amalgamation methods for spatial OLAP and spatial data mining, 6th Int. Symp. on Large Spatial Databases (SSD'99), Hong Kong, pp. 167-187.