Multi-scale Spatial Database Integration

Hande DEMİREL, Turkey and Lothar GRÜNDİG, Germany

Key words: spatial, multi-dimensional, adjustment techniques.

SUMMARY

Due to user requirements, it is very common to encounter situations where multi-scale spatial data is necessary. Data integration is highly desired due to the decreasing long-term costs of obtaining and maintaining data, as well as its beneficial effects on data consistency and accuracy within agencies. Paradoxically, complete and integrated multi-scale spatial databases are very rare. In addition, as user requirements vary considerably, information is also of varying quality. Since an important aim of Geographic Information System (GIS) is the integration of data sources, in particular spatial data, adjustment techniques have proven to be the most effective tool. This is due to the technique's provision of consistent data and the ability to determine accuracy. In order to present the efficiency of adjustment techniques at multi-scale spatial database integration the implementation of non-planar topology, considering the height information is selected. Results are going to be discussed at this study.

Multi-scale Spatial Database Integration

Hande DEMİREL, Turkey and Lothar GRÜNDİG, Germany

1. INTRODUCTION

One of the major current Geographic Information Systems (GIS) problems identified is data acquisition, since organizations can be easily frustrated by the high cost of GIS data acquisition and update. This is not because the information necessary for the agency is not available, but because the information is inadequate for the purpose of the system. Many of the available information is still stored as analogue format, others were isolated within divisions. There was no or very limited data exchange between divisions and departments, even when it was desired. Although there are efforts to eliminate this problem by using interface or middleware, the required efficiency cannot be reached because of redundancy. Many systems were used and maintained in parallel by highway administrations. It was very common to encounter situations where multiple sources of the same information exist. The diverse requirements are stressing agencies to collect the required information by means of various methods, depending upon required generalization level and quality. In addition to these, collected data have generally spatial character. The spatial data acquisition requires the highest efforts and costs, due to the high level necessity of completeness, actuality, correctness and well-defined data structure [BILL,1999] Furthermore, no formal conceptual data model had been designed, including spatial and non-spatial data. Software vendors' proprietary databases were used for storing spatial information. Other information sources, which were unstructured, were then linked as non-spatial information. The conceptual data models were designed in order to respond various user requirements separately. As a result, although there was an enormous data collection effort, the expected efficiency of the system can not be achieved. In spite of such complications, reusability of existing data and integration within the system is essential because of obtaining high costs of digital information. This problem promotes the usage of integrated approach and adjustment techniques, since the efforts required for data integration can be reduced and integration can be automated. The adjustment techniques have proven to be the most effective tool for this purpose, due to the technique's provision of consistent data and the ability to determine accuracy. Due to the integration of methodologies, provision of common basics, decreasing data acquisition efforts and the integration of data sources, integrated approach is accelerating. It takes as input a set of databases (schemas and data instances), and produces as output a single unified description of the reality.[DEVOGELE, 2002] Correlated with the required complexity of the system, this task is becomes complicated, since the system success depends on formally describing the different specialized views of reality. A specialized view of reality is abstraction of real world phenomena, which reflects decisions about features and their relationships. In complex systems, due to the variety of user requirements and applications, different aspects of an object are important. In order to respond these various requirements, information's multi-dimensional spatial character with various levels of abstraction should be reflected into the system.

Topological consistency is an important requirement of multi-dimensional spatial databases. In many GIS data models topology and geometry are combined, although topology is an abstraction of geometry. In spite of some advantages, such as the reduction of total data storage requirements and the increasing of query processing performance, many disadvantages arise with data maintenance. This is especially the case when there are geometrical changes occur in the real world phenomena. Since the geometry and topology are combined, displacements in geometry also affect topology. Although, topology is invariant under position, orientation, transformation, shape and size, data maintenance is required both for topology and geometry after every geometrical displacement. Additionally, the topology needs to be related to all dimensions and diverse applications may require multiple representations of topology. The third issue is non-planar topology. With planar topologies links cannot cross each other without creating an intersection. The crossing links must therefore be split into several individual links. Planar networks have many advantages, principally that they are common and simple. However, this does not reflect the reality for networks where links can cross without creating intersection. In order to diminish the problems virtual nodes were introduced. This necessity emerged in order to differ “real” nodes from “virtual” nodes, which are generated due to the usage of planar topology. However, this solution is inefficient in practice due to the high level of data collection requirement and complications at data maintenance.

Another issue arises due to complex relationships between multi-dimensional objects. Methods need to be designed in order to satisfy user requirements, provide integrity, control the redundancy and certificate the quality. In GIS, due to redundancy, integrity constraints are required wherever geometry interacts. Due to the limitations and requirements of current GIS software for data modeling, many additional integrity constraints are required in order to validate geometry and topology. In order to achieve the required efficiency adjustment techniques can be used. Adjustment techniques is being established as a more standard device in different areas of geodesy. Contraliy at spatial information systems, there exists few examples and the advantages of this technique are under estimated. Adjustment techniques will automate the detection of inconsistencies in multi-layered spatial representations. They will also automate certification of quality and provide integrity.

2. A SAMPLE CASE

The real-world phenomena road is a good example for information’s multi-dimensional spatial character. During this study the concentration was given to road object, however it is obvious that many multi-dimensional objects exists in spatial databases, such as utility lines, facility objects, linear objects, having similar problems. Depending on the required usage, road is multi-dimensionally defined as one-dimensional (linear referencing system), two-dimensional (planar coordinates), three-dimensional (planar coordinates and height information) and four-dimensional (time in the case of dynamic objects). However, in many cases this variety is not fully supported by GIS-Transportation (GIS-T). The most common information referencing method used by highway agencies is the linear referencing system, which is one dimensional. This is based on a one-dimensional specification of the unknown point in terms of direction and distance from a known point. Many other spatial frames are in usage such as numbering systems, addresses, topology, administrative reference systems and road names. In GIS-T roads are defined using two-dimensional reference systems. In order to integrate these various dimensions, typically, geographical location by two-dimensional coordinates is used, and linearly referenced road data is considered as attributive data. However, highway information spatial character is continually changing through new alignments and construction, therefore its reference system is also continuously changing. Therefore, linearly referenced data is badly affected by such geometrical changes, requiring a new referencing for the sections after the modification. Thus, maintenance of this data is clearly necessary. Unfortunately, because of its attributive storage in GIS-T the practical realization is often insufficient. The non planar topology at varying abstractions was not also considered and user assesments were not adequately fulfilled. In order to diminish such problems a multi-dimensional non-planar conceptual data model for a entire highway agency was designed. [DEMİREL, 2002] In the conceptual data model developed, non-planar topology was implemented in order to avoid mentioned problems. In order to implement such non-planarity, use is made of the third dimension namely height information. Part of an example conceptual data model designed of entire highway agency, where topology is designed as non-planar can be found in Figure 2.1.1

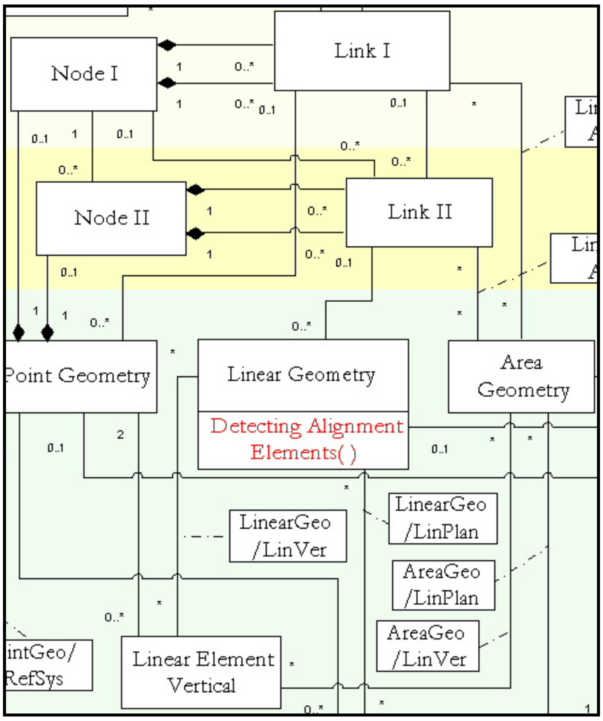

Figure 2.1.1 Part of the Designed Conceptual Data Model

The main elements of topology are *Node* and *Link*. In order to define a link, beginning and ending nodes are required. Secondly, a node can be assigned to many links. These relations are shown in the data model with two *1: 0..** aggregation associations. Associations between *Node* and *Link* are applied to both abstraction levels, using the same aggregated associations. Relationships between two abstraction levels are modeled as follows:

1. *Node I*, being a higher abstraction level may be composed of *Link II* and *Node II.* Between the first level topology element node *(Node I)* and second level topology element node *(Node II*) a *1: 0*..* relationship is assigned. This relationship maps reality adequately because *Node I* may be composed of many *Node II* and every *Node II* is assigned to the first level topological object *Node I.*
2. The relationship between *Node I* and *Link II* is modeled *as 0..1: 0*..*, where *Node I* may be composed of second level links *(Link II)* and a second level link *(Link II*) may be assigned to *Node I***.** Road junctions are examples of such situations.
3. The relationship between *Link I* and *Link II* is modeled as *0..1: 0*..*, where *Link I* may be composed of second level link's *(Link II*) and a second level link *(Link II)* may be assigned to *Link I.* The merging and subsequent separation of divided highways is an example of this.

Using the defined relations, other required information can be extracted. Relationships between topological and geometrical components are described using the following associations. The *Node* object is represented at the geometrical level by a point. *Node I* and *Node II* have a *0..1: 1* relationship with respect to *Point Geometry.* A node must be represented with a point, but a point need not be a node. This relationship is valid for both abstraction levels. *Linear Geometry* has a relationship between the two abstraction levels; *Link I* and *Link II.* Between *Link (Link I* and *Link II*) and *Linear Geometry a 0..1: 1*..* relationship is assigned. A link may be composed of many linear geometry and a linear geometry may be assigned to one link. The *N:M* association between topological element *Link* and *Area Geometry* is realized using association tables *LinkI/AreaGeo* for the first level, and *Link II/AreaGeo* for the second

With the designed objects *Linear Element Vertical* and *Point Geometry,* and *Linear Geometry* object methodology *Detecting Alignment Elements***,** non-planar topology third-dimension is achieved in the conceptual data model. Since the geometrical vertical alignment elements were detected using the *Detecting Alignment Elements*, vertical alignment geometry information is available as "build-in", where these are illustrated in Figure 2.1.2

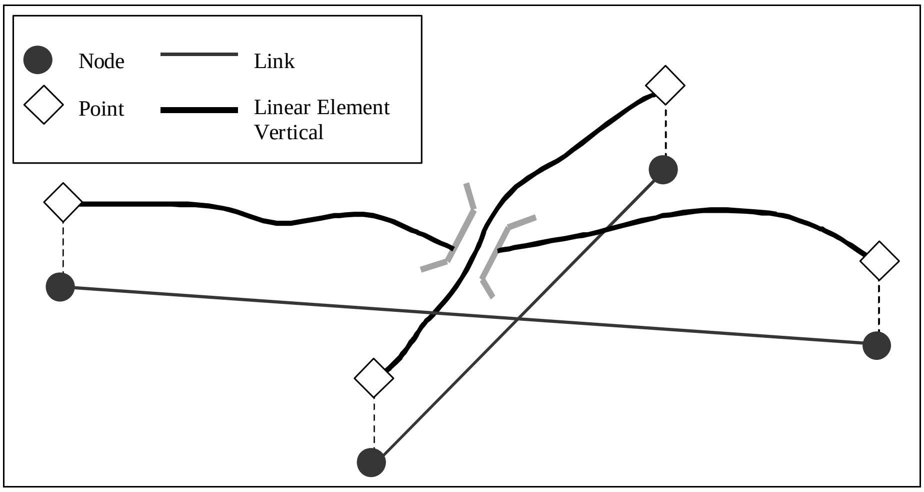

Figure 2.1.2: Non-planar Topology

3. ADJUSTMENT TECHNIQUE APPLIED

During this study, with the use of adjustment techniques many problematic areas identified at GIS-T were solved. As an example, two of them are explained in detail below.

3.1 Detecting Alignment Elements

In the conceptual data model, in order to detect alignment elements the *Detecting Alignment Elements* method of the *Linear Geometry* object is defined. This method is based on the detection of alignment elements realized over significant parameters and elements, with the help of curvature diagram. If the beginning point and beginning tangent angle is known, with an approximation, alignment element parameters and their sequence can be uniquely defined. [GRÜNDIG-I,1988] The curvature diagram is a graphical representation of the curvature (k), where (k) is defined with respect to stationing length (l) as;

$$k = \frac{d\tau}{dl} \tag{3.1.1}$$

Therefore, alignment elements can be identified with simpler functions being; straight lines parallel to axis, straight lines not parallel to axis and quadratic parabola. For each point of the alignment, where the bearing angle is a function of stationing value, the cartesian coordinates (x, y, z) of points of the horizontal (or vertical) alignment results detecting the alignment elements with the use of simple functions integration. Adjacent elements have to fulfil conditions of transition in order to enforce the smoothness of the alignment and of its first derivative. The approximation in the diagram of the first derivative of the alignment, corresponds to a spline analysis using parabolic curves of second order.[GRÜNDIG,1992] The approximation of a sequence of points in the diagram leads to another task. It is necessary to find out the parabolic curve element to which the point has the closest distance. For every point, the bearing angle and the distance is required, where these information can be obtained automatically by means of adjacent points. Since the unavoidable very small distinguishing errors results undesired dispersions in the curvature diagram, during this process generalization effect is used. [GRÜNDIG-II,1988] In order to realize this task adjusted spline analyze with predetermined restrictions is used. Additional constraints for geometrical and driving dynamics are considered in the adjustment model as observations, in order to achieved the optimum result.

Obtained results needed to be optimized. In order to perform this optimization, parameters must be identified. The unique parameters for the straight line, arc and clothoid elements are;

Straight line:	l :	Length of a straight line
Arc:	R, l :	Radius and arc length
Clothoid:	R_P, R_f, l :	Radii of the arcs of the preceding and following element, the clothoid length

Any linear element between point a and point b can be uniquely identified with coordinates x_a, y_a, x_b, y_b and tangents t_a, t_b. With the use of available initial values, the functions of (f) can be linearized.[BAHNDORF,1994] A linear substitute system results which can be solved minimizing a weighted squared sum of residuals of the parameters in a least squares method. As result of analysis:

- Sequence of alignment elements, element type, radius, stationing values
- Coordinate list of alignment main points with approximation values, tangent bearings and stationing values for mentioned points

were obtained. Points sequence, which are representing the road, is required in order to implement this method. This information is available in digitized maps and can be efficiently used. Beginning point and beginning tangent angle can be even obtained from large scaled digitized maps.

3.2 Implementation of Non-Planarity

In order to implement non-planarity, use is made of the third dimension namely height information. Due to various requirements of the highway agency, various height information sources exist. Examples of such sources are;

- The road gradient and road inclination values
- Geometrical design regulations, for example minimum overpass height, driving dynamics and safety regulations
- Digital Terrain Models (DTM)

In order to obtain non-planarity, for the entire highway network, concrete height information for geometrical elements is required. These are only fully available in DTM among the introduced sources. However, several problems were apparent. Firstly, the required accuracy in applications which need non-planar topology, such as Intelligent Transportation Systems (ITS) and freight management, can not be achieved economically. Because DTM accuracy is tightly correlated with economical aspects, depending on the data collection method and map scale. Costs increase as the provided accuracy increases. Additionally, having high accuracy DTM data will not alone fulfill this requirement. This is because DTM data does not match with road structures such as; bridges, tunnels and overpasses in a one-to-one manner.

In order to solve this problem, other data sources, such as; geometrical design regulations, driving dynamics and traffic safety regulations, needs to be introduced to the system. However, in this case the solution is not unique, leading to redundancy. By means of adjustment techniques, the required accuracy can be achieved and the redundancy can be controlled.

The adjustment theory is an established optimization technique used to determine unknown parameters based on given observations. It provides a straightforward solution to the described problems. The aim of least squares adjustment is to optimize the solution of a functional model by minimizing the residuals of the observations.

$$\sum v_i^2 \; P_i = \min \qquad (3.2.1)$$

where, v are the residuals and **P** is the weight matrix containing values corresponding to the observation accuracy.

The unknown parameters $\bar{x}$ can be solved according to the following equation:

$$\bar{x} = (A^T\ P\ A)^{-1} A^T\ P\,(\,l - f\,(x_0)\,) \qquad (3.2.2)$$

where A is the Jacobean matrix of the function derivatives with respect to the unknowns, l are the observations and $f\,(x_0)$ is the value of the function calculated with approximate values. This optimization problem can be solved in one of two ways; direct and indirect. The direct approach is introduced into the system using conditional equations. The indirect approach is generally preferred due to its better suitability for computation and error estimation. With the indirect approach, two options are available;

- Introduce conditional equations between the unknowns.
- Enforce specific observations as being more accurate in the stochastic model.

Since conditional equations produce large normal equation systems, the second approach is preferable.

In order to help clarify the proposed second solution approach, an example illustrating the interpolated DTM height information and one of the constraints, minimum overpass height, is presented. This information is shown in Figure 3.2.1.

Figure 3.2.1: One of the Sample Constraints, Overpass-Height

The unknowns in this example are the relative height differences ($\Delta\, h$)

$$\Delta\, h = H_A - H_B \qquad (3.2.3)$$

The interpolated height observations, representing the relative height differences between points along a road, can be defined in the functional model as;

$$\Delta\, h + v_{\Delta h} = H_A - H_B \qquad (3.2.4)$$

Constraints, in this case overpass height, is introduced into the system, using the same functional model;

$$\Delta\, h + v_{\Delta h} = H_U - H_L = 4.70\ m \qquad (3.2.5)$$

In the stochastic model, each observation's corresponding weight is stored in the diagonal elements of **P** matrix as:

$$P = \frac{\sigma_0^2}{\sigma_{\Delta h}^2} \qquad (3.2.6)$$

For this example, the standard deviation of point heights obtained from the interpolated DTM is assumed to be $\sigma_{\Delta h} = \pm\, 5\ m$, and the observation variance is $\sigma_0^2 = \pm\, 1\, m$. The overpass-height observation is introduced with a standard deviation of $\sigma_{\Delta h} = \pm 1\,cm$. These introduced source, which is acquired more accurately, define conditions in the stochastic model. Since $\sum v_i^2\ P_i$ must be a minimum and every observation must completely fulfill the conditions, the introduced overpass-height constraint enforces the model in order to fulfill its condition. Since the standard deviation information is “fixed”, other parameters must change, including the observations. This process is performed iteratively, until $\sum v_i^2\ P_i$ is minimum and all conditions are fulfilled. Consequently, when the proposed method is applied to a system where the height differences are interpolated from a DTM of lower accuracy, the non-planar topology can be optimized for the entire highway network. Therefore, high accuracy expectations are fulfilled at low cost. However, with this approach there is a risk of introducing very “strong” constraints, which results in undesired deformations of other observations in the system. This is due to, an adjustment approach allows for a change of all parameters while simultaneously enforcing constraints together with the tolerance of the constraining points. This problem can be solved, by loosening the “strong” constraints until the appropriate solution is achieved.

4. CONCLUSION

With the prevalence of GIS usage and the resulting raise in data sharing issues, spatial data integration efforts are increasing worldwide. In order to materialize the requirements of the user, to model the reality in an appropriate manner and to increase the efficiency of the system, multi-dimensional spatial information systems are obligatory. In order to perform this critical and complex task, adjustment techniques provides adequate and cost-optimized tools. This optimization is achieved by means of consistent data produced, automated error detection and certification of accuracy and reliability. Since adjustment techniques offers standardized and mature solutions, which are previously applied in various areas, the proposed solution is a general and global solution.

REFERENCES

Bill, R.,Grundlagen der Geo-Informationssysteme, Band I Analysen, Anwendungen und neue Entwicklungen Wichmann, ISBN 3-87907-326-0, Heidelberg, 1999, pg 167-172

Devogele, T., Spaccapietra, S., On Spatial Database Integration, http://lbdwww.epfl.ch, (retreived in 2002)

Demirel, H, An Integrated Approach to the Conceptual Data Modeling of an Entire Highway Agency Geographic Information System (GIS), Verlag der Bayerischen Akademie der Wissenschaften in Kommission bei der C.H.Beck’schen Verlagsbuchhandlung München, ISBN: 3-7696-9593-3, 2002, Germany

Gründig L., Datenverwaltungskonzepte für typiche Aufgaben aus der Ingenieurgeodäsie, Deutsche Geodätische Kommission, Reihe C.Nr.336, 1988,pg.38-43

Gründig L., Bahndorf J., Singer P., The Design and Management of the three Dimensional Alignment of Rails and Switches using RAILCAT, Comp' Rail 92 Proceedings of an International Symposium, Washington, 1992

Gründig L., Bahndorf J.,Ein rechnergestütztes Entwurfsystem zur drei dimensionalen Trassierung, X. Internationaler Kurs für Ingenieurvermessung, München, 1988

Bahndorf J., Gielsdorf F., Gründig L. , The Design and Management of the Three Dimensional Alignment of Rails and Switches, FIG XX. International Congress, Melbourne, Australia, 1994

CONTACTS

Dr. Hande Demirel
Istanbul Technical University,
Civil Engineering Faculty,
Division of Photogrammetry
Ayazaga Campus
80626 Maslak
Istanbul
TURKEY
Tel. + 90 212 285 6110
Fax + 90 212 285 6587
E-Mail: hdemirel@ins.itu.edu.tr, hande.demirel@itu.edu.tr
Web site: http://www.ins.itu.edu.tr/jeodezi/fotog/hdemirel/

Prof. Dr. Lothar Gründig
Berlin Technical University
Fachgebiet für Geodäsie und Ausgleichungsrechnung
Sekretariat H20
Straße des 17. Juni 135
D-10623, Berlin
GERMANY
Tel. + 49 030 31 42 23 75
Fax + 49 030 31 42 11 19
E-Mail: gruendig@inge3.bv.tu- berlin.de
Web site: http://www.survey.tu-berlin.de/index_1.html