Spatial Data Structures UnderThe Fourth Dimensional Time Axiom

by

Harold G. Campbell, Ph.D.
Associate Professor Of Computer Information Systems
Humboldt State University

Introduction

The recent explosion of microcomputer based geographic information systems (GIS) and spatial information systems (SIS) packages have brought about a myriad of applicational benefits to end-users engaged in a variety of technical and general industries. These types of software products now reside on a diverse crosssection of hardware platforms and are supported by nearly all of the common operating system architectures. GIS based applications can be found to include such disciplines as; urban planning, management of natural resources, medicine, business and marketing, command/control of military forces, and even the mapping of space. Several of the more well respected journals have indicated that these types of systems are likely to continue to gain in popularity because of their ease of use and ability to model real-world phenomena and will most likely continue to capture substantially larger shares of the personal computer software market over the next several years.

Theoretical Axioms

Of the multitude of theoretical axioms which support the function and utilization of these types of software, perhaps none are more important than those derived from Einstein's special theory of relativity and Minkowski's axiom of four dimensional space. Although most end-users would not consciously draw a linkage between Einstein/Minkowski and GIS, many of the routines they employ daily as a matter of interface with such systems display a direct lineage to these rather complex theoretical postulates. As prescribed by these theories, space (and all that exists or transpires within space) does so relative to a four dimensional continuum. Three dimensions occupy the standard Euclidean space coordinates (x,y,z) and one dimension is depicted as the time (t) coordinate during which the event happened. To conform to the natural laws of physics an entity must exist and occupy a finite space which can be represented relative to either single or multiple coordinate systems and such an entity's existence must have transpired within a similar partition of time. Prior to the advent of the theory of relativity, time was considered to possess structural independency. Now however, the time coordinate plays exactly the same role (under special relativity) as do the space coordinates and all assume consistant mathematical properties. This we find important in not only describing physical phenomena, but (using GIS/SIS) modeling the universe and such phenomena. For purposes of this paper, it is necessary to focus upon the importance which Einstein gave to the space-time coordinate system that support the special theory of relativity (x,y,z,t) and the Lorentz transformation (x',y',z',t') rather than that system used to support the general theory of relativity (x1,x2,x3,x4). The reason for this is the presumed absence (hypothetically) of any contaminating gravitational fields within the spatial systems being replicated by most end-users.

The Current Evolution

As presently evolved, most of the data structures used in support of GIS/SIS concentrate on the successful blend of spatial information (normally generated under the standard spherical coordinate system or converted from a Cartesian system) with selected attribute information that have relevance to an entity which occupies a specific point or polygonal boundary. A large number of GIS/SIS software producers have chosen to focus soley upon the X and Y axis for display and representation of digital line graph (DLG) imagery and a few now incorporate a Z axial position to support digital elevation model (DEM) requirements. Once established as an acceptable set of recognized coordinates within a map file or boundary file, attribute data contained within a pointfile/database are geocoded to the spatial positions by matching a selected attribute value contained within a record against the map or boundary file and held there by appending the corresponding coordinate value to this record. Despite the fact that these types of software packages have radically transformed the manner in which data are retrieved and manipulated by end-users, and have most certainly enhanced general data utility, specific deficiencies regarding the validity/reliability of attribute and spatial data can be identified which could loom as a major source of future problems for users. These deficiencies are not necessarily inherent to the limitations of geographic or spatial information systems technology, but rather are a consequence of the data structures currently used to support these systems.

The Problem

What is curiously absent from the data structures used by most end-users is a field which identifies the time element (t) at which an association is valid. Despite the fact that virtually all of the currently marketed GIS/SIS packages allow for the inclusion of a fourth spatial variable, seldom in practice, is "time" ever used to assure data validity and reliability. It is my suspicion that the absence of a time axiom from such databases is mostly a matter of three dimensional familiarity by those who use these products rather than an inability to recognize the importance of such a variable in assuring accuracy. Since multidimensional perspectives tend to confound interpretation, then by limiting spatial elements to only two or three axial planes, one can simplify the problems associated with interpreting

perspective. The unfortunate consequence of this practice is likely to be however, that a severe loss of accuracy will occur as databases become protracted or as dimensional perspectives are extended to encompass multiple coordinate systems.

Spatial/Attribute Integrity

Before it is possible to gain a full understanding of the functional aspects of this situation, it is necessary to first examine these dimensional perspectives issues and their potential effects. In Figure I, an identifiable point with a distinct degree of singularity (the earth) has been depicted and is enveloped by an expanding sphere which represents all three standard axial planes, but which also represents the time dimension. As the enveloping sphere expands to (t+1) the location signified by the X,Y, and Z values holds constant (assumming no motion) but at a new time along the progressing continuium. As this time continium progresses, locational positioning remains static, but perhaps with a change in attribute values associated with a particular point or with an entirely new attribute associated with this point, but at a new time. Since data to support GIS/SIS can be derived from either network, hierarchical, or relational database models and in consideration of the fact that GIS/SIS is merely a front-end interface with a database engine, it becomes necessary to contend with the value fluctuations or the phenomena of "attribute genesis" within the data structure to accurately model what is actually occuring. That is, the alteration in value or evolution of an entirely new attribute associated with such a point that did not exist at (t) but now exists at (t+1). Since it is not only theoretically possible for a single point to possess multiple attributes at the same time, but is quite common, one must contend with this fact within the data structures used to model the universe and the only dimensional marker available to measure such an occurrence

Figure I
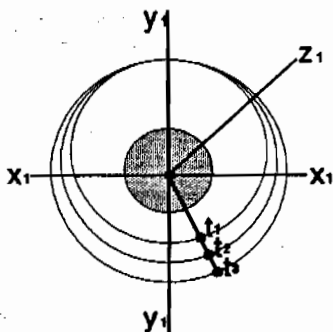
is "time". Without such a marker, it becomes impossible to determine when an attribute winks into existance or ceases to exist. Since the majority of GIS/SIS efforts attempt to portray a situation at a past moment in time or forecast future events, it is therefore critical to such analyses to possess the time marker within the data structure itself as a secondary or terciary key. Without such a key field, events cannot be sliced into manageable components and modeling of the real-world cannot occur.

The problems associated with not defining the time variable within the data structure become even more pronounced when

multiple coordinate systems are used and when rotational movement and velocity are involved. In Figure II, the spatial location of a point is easily distinguished by both coordinate systems and subsequently its attributes as described within the database. As a full rotation occurs however, and if the forward velocity of the object is insufficient to alter its relative position to the second coordinate system, then the point will appear to have realigned itself to the second coordinate system but perhaps with a change in value for a specific attribute or with the existance of an entirely new attribute. Again, the only effective

Figure II
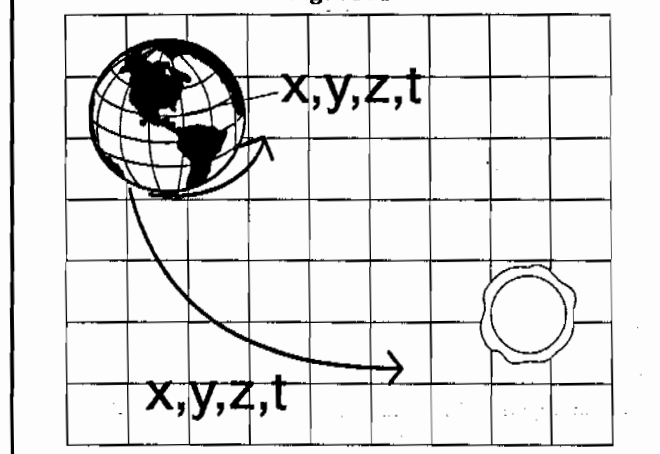
methodology to assure the valid modeling of such an event appears to be by designing the GIS/SIS inclusive of a "time" variable so as to provide for incremental progression/regression. The more coordinate systems you add to the equation, the more pronounced the problem becomes, and the greater the reliance upon time to separate the events. When exponential levels of motion and relativity are encountered (such as three dimensional movement upon the surface of the earth, combined with the earth's rotation about its axis, along with the orbital movement of the earth around the sun, the sun around the galaxy, and the galaxy's movement away from the center of the big-bang) the need for "time" as a marker becomes imparative. Non-geographic based applications such as those employed to model production systems or systems designed to represent and control processes are especially vulnerable to multiple perspective orientation problems. Without distinct "time" markers within these types of modeling systems, it is virtually impossible to monitor flow and even more impossible to understand cause and effect issues or lag-lead phenomena.

Component File Layering

Admittedly, a strong argument can be made that the practice of component file layering (the process of overlaying points and boundaries on top of one another based on a time variant) partially satisfies the need for time measurement. However, as currently practiced, layering focuses on shifting boundary polygons according to a "time sequence" or individual slice in time and cannot easily exclude point files or attributes which did not exist at the time frame that all of the layers represent. In other words, invalid attributes may simply show through aggregated time layers and become spurious or misleading. Aggregated dimensional mechanisms which incorporate spatial positioning along with attribute or entity values require an exclusionary view in order to gain an accurate and comprehensive perspective. Exclusion of a formalized

time variable from the database itself does not serve to assure required data integrity in these cases and certainly necessitates the multiple processing of data or physical separation of databases. Such practices cause further data management complications when attempting to "generate" boundaries or thematically represent data from database information in an interative mode such as in the case of weather or brain activity mapping. The absence of a time variable within these types of databases makes it impossible to search for temperature or pressure readings at a specific time and adjust polygonal boundaries interactively as time advances. Instead, sequential maps or layers must be drawn based on non relational database constructs requiring excessive data handling and allowing for the potential contamination caused by the aggregated effects cited above.

A Sample Data Structure

Figure III below illustrates a sample data structure which incorporates the time variable along with two attributes, and a "Location" field, as well as the optimum spatial elements. Under such a structure, it is possible to distinguish that the hypothetical values for Attribute (1) and Attribute (2) are directly relative to the defined locations as of (t=01/01/91). In this case, time has become a secondary key field and acts in conjunction with "Location" to assure data validity. If the "Time" field were not present in this case, then it would be impossible to distinguish when the attribute values became legitimate. Without time, the data set could mistakenly include attribute values that were relative to the specific geographic location, but were perhaps not necessarily relative to the same entity which occupied that specific location. Record three has been added to illustrate the problem caused by aggregation.

Figure III

Entry	Att1	Att2	Location	Xcoord	Ycoord	Zcoord	Time
Rec 1	AAA	BBB	12 A Street	39.123	123.123	1000	01/01/91
Rec 2	EEE	FFP	13 A Street	40.123	124.123	1000	01/01/91
Rec 3	KKK	LLL	12 A Street	39.123	123.123	1000	01/02/91

Without the time field, under aggregation, it becomes very difficult to determine changes within the attribute value fields further complicating the process of assuring data validity.

As mentioned earlier these types of data structuring problems can become even more acute as time becomes protracted and as additional attributes are added to the database. Figure IV depicts such a situation where a third attribute is added to the database at (t+1). As depicted, the values for Record 1 (Att1 and Att2) remain unchanged at the new time, and Att3 is valued at CCC. For "Record2" however, Att3 is an illegitimate field and may well be represented by a spurious value unless a time variable is present within the database to act as a secondary key or to restrict invalid data associations. Simply assigning a "null" value to such fields cannot adequately guard against contamination within a GIS/SIS setting because of the potential effects caused by aggregation.

Figure IV

Entry	Att1	Att2	Att3	Location	Xcoord	Ycoord	Zcoord	Time
Rec 1	AAA	BBB	CCC	12 A Street	39.123	123.123	1000	01/02/91
Rec 2	EEE	FFF	N/A	13 A Street	40.123	124.123	1000	01/02/91
Rec 3	HHH	III	JJJ	14 A Street	41.123	125.123	500	01/02/91

Conclusion

Review of these data structures appears straight forward enough when presented within this rather stepwise explanatory process. But if you ponder the difficulty which would occur, and most certainly the inaccuracy which would result, if the "time" values were not presented as part of these analyses, it becomes clear that the functionality of these spatial information systems is strongly reliant upon Einstein's special relativity theory and Minkowski's time axiom. Failure to consider the time axiom within the dataset used to support GIS/SIS in essence violates the natural laws identified and expressed by these men and could result in gross misinterpretation of information and correspondingly inaccurate conclusions about the phenomena under study. It is critical to remember that under the special theory of relativity, time plays exactly the same role as do the spatial coordinates and the exclusion of time from a GIS/SIS database subsequently incapacitates these computer-based modeling efforts.

Finally it should be pointed out that the methods of time quantitification and representation within the data structure are also as important to the successful utilization of GIS/SIS as are the exact delineation of surface structures/features and entity/attribute data. Depending upon the particular endeavor, time might need to be quantified into excessively small components to signify the date, hour, minute, and second of association. Even smaller increments are required for some of the more advanced applications.

A solid rule to assure the successful design, utilization, and application of GIS/SIS systems is simply *"to model the real-world, you must reconstruct the real-world, and as Einstein and Minkowski demonstrated, there is always time."*

About The Author:
Dr. Harold G. Campbell is an Associate Professor of Computer Information Systems with Humboldt State University, California. He received his Ph.D. from Claremont University in 1983 and has specialized within the disciplines of mathematical analysis and spatial information systems technology.