A RELATIONAL DATABASE MODEL FOR AN AVL SYSTEM AND AN EXPERT SYSTEM FOR OPTIMAL ROUTE SELECTION

Hassan A. Karimi, Edward J. Krakiwsky,
Clyde Harris, Guy Craig, and Rob Goss
Department of Surveying Engineering
The University of Calgary
2500 University Drive N.W.
Calgary, Alberta, Canada T2N 1N4

ABSTRACT

During the course of prototyping an Automatic Vehicle Location (AVL) system it was discovered that more than one database is needed, and that the map database (city, province, state ..., etc.) was conceived of being the primary database. The use of a microcomputer along with a real-time requirement are the two dominant constraints on an AVL map database which differentiates it from other types of map databases. Specifically, real-time response has an impact on the nature of database communication, and use of a microcomputer limits the primary and secondary storage capacity. After creating the infological (real-life) model it was concluded that a relational database was needed for the datalogical (computer-oriented) model. A knowledge base which is derived from the map database, is envisaged as the prime source of information for the design and implementation of an expert system for the optimal route selection.

THE UNIVERSITY OF CALGARY'S AVL SYSTEM (AVL 2000)

AVL Systems

An Automatic Vehicle Location (AVL) system is mostly referred to as an assembly of technologies and equipment that permits centralized and automatic determination, display, and control of the position and movement of a vehicle throughout an area [Skomal 1981]. A number of investigators, among which are automobile manufacturers, have already taken up the challenge to design and develop AVL systems. Although these investigations have common objectives, they are different in their approach and techniques. For example, AVL systems employ different positioning techniques (e.g., dead-reckoning, radio communication, or satellite positioning), depending upon the positioning accuracy needed, the cost, etc.

Several alternatives exist when developing an AVL system. Choosing which alternative should be used depends on the requirements. Some applications of AVL systems include emergency, fleet management, delivery trucks, transit buses, taxicabs, and couriers. For some of the applications, a dispatch is necessary; for the others it is an option, hence the requirements dictate the configuration. Clearly, one AVL system could not satisfy the requirements of all different applications, thus a separate AVL system is needed (special-purpose AVL systems).

The hardware component of an AVL system includes a positioning facility, a computer, input and output facilities, and a storage facility. The software component of an AVL system is mainly composed of a control program, databases, interfaces, and input files.

Prototype

The development of an AVL system with the extension of its application in surveying and mapping, has been initiated at The University of Calgary [Krakiwsky et al. 1987] - AVL 2000. From the start, the overall design criteria considered the present needs and the future advancement. The objective of the project was to build an in-vehicle real-time system which utilizes a GPS receiver for positioning and a digital map for position superimposition. Also, an optimal route algorithm was to be built into the software component. That is, by choosing a destination point, the best route - based on some criterion such as time, distance, ... etc. - between the source and this point is computed in real-time. To realize and solve the practical problems associated with the above objective and to pilot future trends, the AVL 2000 system concept was prototyped.

A microcomputer (Macplus with 1 megabyte RAM), a 20 megabyte hard disk, an input device (keyboard), an output device (CRT), and a GPS receiver (Trimble) make up the hardware component. The choice of the hardware components was not considered crucial as it was mostly the software component which had to be prototyped. A control program, map database, optimal route database, and input coming from the receiver comprise the software component. Some of its functions include map display, highlighting the optimal route, and indicating the moving vehicle position on the map. Figure 1 illustrates the configuration of the AVL 2000 prototype.

First and Second Generation Systems

The design of the first generation AVL 2000 system is presently being developed with specific applications in mind. The hardware component of the first generation is to be composed of a microprocessor, CD-ROM, and a voice synthesizer for both input and output. Some AI concepts will be part of the first generation software component.

Clearly, the prototype is microcomputer-based while the first generation is a microprocessor-based (a microprocessor totally devoted to the AVL tasks). The second generation system will be a VLSI-based (a VLSI chip designed to perform all the required AVL tasks).

MAP DATABASE

One of the key elements in any AVL system is the digital route map. Regardless of the application area, the

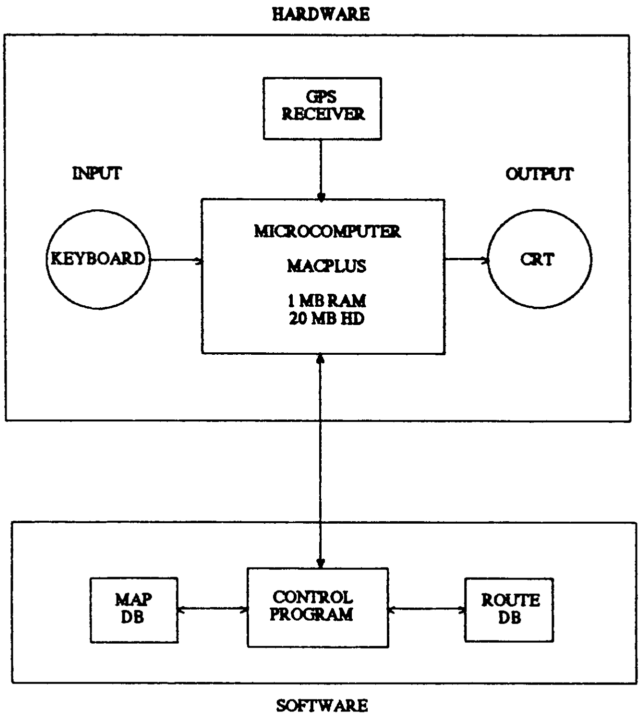

Figure 1. The AVL 2000 prototype configuration

spatial data of an area's road network along with the spatial topology organized and managed in a database are required. This map database was even needed at the very early stages of prototyping.

<u>Requirements</u>

Since the AVL 2000 system is a microcomputer-processor based system, has a real-time system requirement, and has limited storage, the nature of the map database is affected significantly. The requirement of being a real-time system has an impact on the response of database communication, while limited primary and secondary storage limits the quantity of spatial data in the system.

Sources

Two plausible sources of a digital street map are perceived if it is not already available in the form usable in an AVL system. The first one involves digitizing map sheets of various scales; the second one involves adaptation from any existing digital map, which may have been originally developed for other purposes. The answer to the question of which method should be used will depend on the availability of the existing digital maps. For example, the ETAK [Honey et al. 1986] database is derived from a combination of U.S. Census Bureau Dual Incidence, Matrix Encoded (DIME) files, U.S.G.S. 7-1/2 minute quadrangles, aerial photographs, and local source material as needed.

Obtaining a digital map database was perceived as being the main bottleneck during prototyping. Digitization was ruled out as it was more time consuming. A digital highway network of the province was the only suitable source in our operating area. The provincial file (a flat file) had to be transformed into the form of a database usable in the AVL 2000 system.

Storage

AVL systems to date have used three different methods of map storage: photographic, digital image and digital encoded. Honda's Gyrocator and Omni Devices' navigator both use photographic map storage [Cooke 1985]. Chrysler's Stealth uses digital images (video disk images) of paper maps in its "CLASS" (Chrysler Laser Atlas Satellite System) [Cooke 1985]. A digital encoded map is a data file describing the road networks by topology using vector format. This method is more flexible and more compact than the first two methods. The ETAK [Honey et al. 1986] navigation system uses digital encoded map storage. This method more easily provides flexibility in windowing and zooming than the other two. Also, this method is more appropriate for the computations required for optimal route selection algorithms [Cooke 1985].

In the prototype, the highway map database was digitally encoded. Some of the reasons for this include flexibility, storage compactness, map display, and selection of destination by street address or intersection. Of these, the compactness of the database storage is of particular importance. Even when digitally encoded, the quantity of spatial data is usually very large; for example, the road network of the San Francisco Bay area requires between 7 and 10 megabytes of data storage space [Zavoli et al. 1985].

Data Modelling

Once the spatial data of a road network are gathered and stored, they have to be organized and managed along with their spatial relations in the form of a database system. Different approaches for formally representing relationships in a defined data structure give rise to various data models - most notable are the hierarchical, network, and relational models [Everest 1986]. In the network data model, many-to-many relationships can be represented. The hierarchical data model, which actually is a special case of the network data model, represents one-to-many relationships. In these two data models, all relationships between entities must be explicitly defined for the physical access paths. In contrast to these two data models, the relational data model excludes physical access path information. All relationships are implicit in explicitly defined attributes in the entry types.

In this project, after some investigation, it was realized that any of these data modelling approaches could be used, but implicitly defining relationships would provide more flexibility in structuring and future updates. Consequently, the relational data model was implemented. Moreover, as this map database may be used by some other modulas, such as optimal route determination, the relational model would provide all these various views more easily and in a shorter time. The AVL real-time characteristic was considered to be the main constraint in implementing a database in the system. It was decided that a polygon-based relational database would satisfy this. In a polygon-based database, a given digital road map is divided into polygons, and each polygon is used as a reference entity. Figure 2 illustrates the AVL 2000 polygon-based concept (the entities and their relationships) for road networks. Not only will this method support the necessary real-time response (since at any time, only a small portion of a large digital map is accessed and used) but also this makes it easier to implement the optimal route determination algorithms. In this database, one relation (file) was designated as the polygon directory which contains the most immediate information such as polygon ID, number of nodes in each polygon boundary, etc. as its attributes.

Different applications need the same digital spatial data of the road map, but each would view it differently. It is suggested here that a polygon-based relational database can more easily support these different views of spatial data. This can be done by means of defining polygon boundaries for each separate application. This is to say that the original route map can be divided into polygons many times, and each time the criterion for boundaries can be related to the requirements of a different application area. Doing this in developing an AVL system will optimize the flexibility in views, real-time response, and user-friendliness of the system. On the other hand, applying the same polygon boundaries for all different applications, would allow the system to be optimized for only a few cases.

OPTIMAL ROUTE DETERMINATION

Optimal route determination is the computation of the "best" route between the source node and a selected destination node. "Best" could mean shortest, fastest, safest, etc. depending upon some criterion related to the underlying application. For example, for an emergency

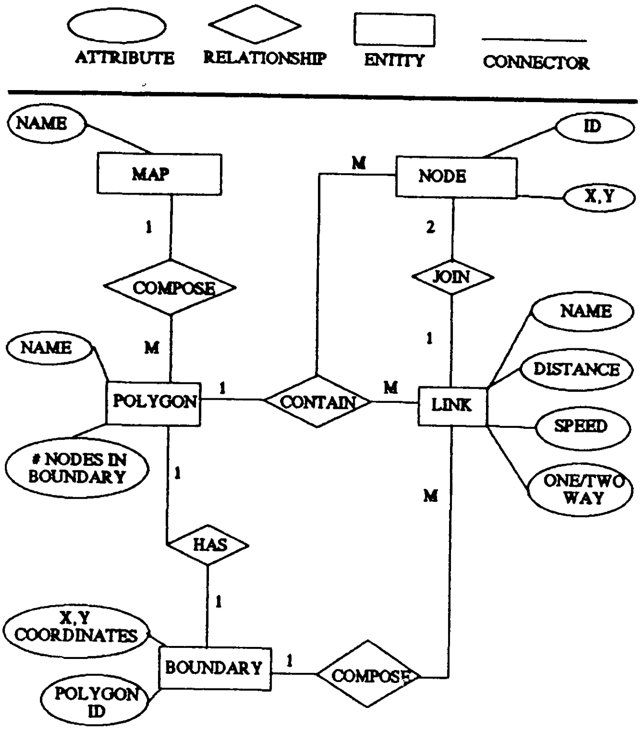

Figure 2. Elements of Entity-Relationship Model (ERM) and a geometric ERM for the AVL 2000 system.

ambulance driver, "best" usually means the fastest route, whereas for a tourist driving in a car, it could mean a route with more scenery.

Mathematically, a collection of a set of nodes and a set of links is called a network. Types of networks, each for a certain application, range from the simplest form of deterministic non-time-dependent travel times networks [e.g., Dijkstra 1959] to probabilistic time-dependent travel times networks [e.g., Hall 1986]. Algorithms have been developed in the past to determine the optimal path between any pair of nodes in such networks. These algorithms vary according to the type of network and type of application they were designed to handle. However, optimal route determination algorithms are known to be NP-complete [Sedgewick 1983].

In the AVL 2000 prototype, a deterministic non-time-dependent travel times network was assumed. Deterministic networks are networks whose links are associated with a specific known weight. Such a weight might include the length of a link or the link travel time. Non-time dependent networks are networks whose link weights do not deviate over time. For the prototype, three classes of weights were considered: (a) distance (travel length), (b) time (travel time), and (c) cost (travel cost). Of the three, only distance is an independent weight, the other two are functions of distance. For simplicity, distance (travel length) was used in the prototype. When the weight is distance, the optimal route is referred to as shortest path. Several algorithms have been developed for the standard node to node shortest path problem. Of these, Dijkstra's [1959] algorithm has been shown to be the most computationally efficient. Its computational time is proportional to n^2 ($O(n^2)$, where n is the number of nodes in the network. One of its shortcomings is that it cannot handle negative weight, but this is not considered as a problem in the AVL 2000 prototype (no negative distance).

Patch-Quilt Concept

It was mentioned that the computation time for this algorithm ($O(n^2)$) is better than some of the alternatives. However, even this is not adequate in a real-time system, especially when a large network is used. To overcome this problem, a "Patch-Quilt" concept was developed. It simply means that for those optimal route selections that require a significant amount of time, both Dijkstra's algorithm and a look-up table are used. The look-up table contains the optimal routes between any pair of polygons in the network. The look-up table entities were computed a priori using a VAX/VMS system. To compute the shortest path between any pair of polygons, the many-to-many nodes shortest-path algorithms must be used. This is because each polygon has many entrance-exit nodes. Such algorithms have been developed by a number of investigators such as Floyd [1962] and Dantzig [1966]. These algorithms are more efficient than almost any one-to-one node algorithm. This is generally true because the application of a one-to-one algorithm for a many-to-many case requires a large number of iterations.

If the destination node occurs in the same polygon as the source node, Dijkstra's algorithm will be used. This can be done in real-time without much overhead since each polygon is a fragment of a large network. On the other hand if the source and the destination nodes occur in separate polygons, the Patch-Quilt concept will be used. To accomplish this, three pieces of the final optimal route between the source and the destination are computed individually. These three are:

(a) optimal route between the pair of polygons;
(b) optimal route between the source node and the entrance-exit node marked by (a) within the source polygon; and
(c) optimal route between the destination node and the entrance-exit node marked by (a) within the destination polygon.

Optimal route between the pair of polygons is retrieved from the look-up table. This route joins one of the entrance-exit nodes of the source polygon to one of the entrance-exit nodes of the destination polygon. Then Dijkstra's algorithm is used in real-time for part (b) and (c). Finally, these three pieces will result in the required optimal route between the source and the destination.

EXPERT SYSTEMS IN AVL

During prototyping, several problem domains of AVL systems applicable to expert systems development were discovered. These include the AVL system management software, database management, database interface, and optimal route determination. In this paper, the concept of an expert system for optimal route determination is discussed, but before doing so, let us give briefly the definition of an expert system used herein.

Expert systems are computer systems that think and reason as an expert would in a particular domain, and solve real-world problems using a computer model of expert human reasoning, reaching the same conclusions that the human expert would reach if faced with a comparable problem [Weiss et al. 1984]. In contrast to a conventional program, an expert system is believed to have self-knowledge: knowledge about its own operation and structure [Hayes-Roth et al 1983]. A system often confused with an expert system is a knowledge system. Expert systems perform as human experts do whereas knowledge systems perform simpler tasks that normally require human intelligence.

In the prototype, an algorithmic approach was taken as the basis for optimal route implementation. It is seen that this technique may not be sufficient for some applications, as for the solution it only requires parameters whose values are certain and obtained a priori. Certain applications may require some other parameters that cannot be obtained a priori and whose values are time-dependent. These parameters, which can be called real-time parameters, include time of the day, season of the year, weather conditions, accident information, detours, road condition, traffic flow, visibility, and purpose of trip. Some of their values, which can be input to the system interactively, are uncertain. This would result in uncertain inferences.

Clearly, introducing real-time parameters to the system would make the task of reasoning and decision-making difficult enough to require expertise. For this, an expert system specialized in optimal route determination is developed for the first generation system. The domain-specific expert must find the best routes in a network in real-time using all real-time information. An expert who does this can be called a "route finder consultant".

The first generation system will use the algorithmic methodology blended with heuristic knowledge. This is one of the main characteristics of a true expert system. A heuristic is a rule of thumb, a trick, strategy, simplification, or other method that aids in the solution of complex problems. Heuristic generally reduces the size of the space in which one needs to search for solutions to the problem at hand.

The rules and facts associated with both the real-time parameters and the map database will be used in the construction of the system's knowledge base. A knowledge base - a collection of rules and facts - is the main source of information for the inference engine (reasoning strategy) of an expert system.

Further, for the treatment of uncertainty in this expert system, one of the several methods which have been developed to take the uncertainty into account will be considered (e.g. Bayesian method, or the concept of "fuzzy set" originated by Zadeh [1978]).

CONCLUSION

Specific application requirements should be the basis of the design and the development of an AVL system. A digital encoded map is recognized as the best map storage technique. This provides and supports many common and uncommon features required by most applications. Of the three well-known data modelling approaches, the relational model provides and supports more easily the real-time response, compact data storage, and different views of the spatial data. An algorithmic approach was taken for the optimal route determination in the AVL 2000 prototype. It was realized that this would involve a significant amount of overhead in a real-time system. For this, a "Patch-Quilt" concept using a priori polygon to polygon shortest-path computation was developed. For the first generation, it is intended to include some AI techniques and tools, such as expert systems, into the system. Among problem domains of an AVL system applicable to expert systems development, optimal route determination is of particular importance. For this, real-time parameters whose values are input interactively along with some information derived from the map database will be the basis of defining rules and facts to form a knowledge base.

ACKNOWLEDGEMENT

This research was conducted under a Natural Sciences and Engineering Research Council of Canada grant.

REFERENCES

Cooke, D.F., Vehicle Navigation Appliances. Proceedings on Digital Representations of Spatial Knowledge, AUTO-CARTO 7, Washington, D.C., March 11-14 (1985).

Dantzig, G.B., All Shortest Routes in a Graph. Operation Research House, Stanford University, Technical Report 66-3, November (1966); also in Theorie des Graphs, pp 91-92, Proceedings of the International Symposium, Rome, Italy, July (1966), Published by Dunod, Paris.

Dijkstra, E.W., A Note on Two Problems on Correxiom with Graphs, Numerische Mathematik 1, 269-271 (1959).

Everest, G.C., Database Management: Objectives, System Functions, and Administration, McGraw-Hill, Inc. (1986).

Floyd, R.W., Algorithm 97, Shortest Path, Comm. ACM 5, 345 (1962).

Hall, R.W., The Fastest Path through a Network with Random Time-Dependent Travel Times, Transportation Science, Vol. 20, No. 3, August (1980).

Hayes-Roth, F., Waterman, D.A. and Lenat, D.B., Building Expert Systems, Addison-Wesley Publishing Company, Inc. (1983).

Honey, S.K., White, M.S. and Zavoli, W.B., Extending Low Cost Land Navigation into Systems Information Distribution and Control. IEEE Symposium on Position Location and Navigation, Las Vegas, Nevada, November 4-7 (1986).

Krakiwsky, E.J., Karimi, H.A., Harris, C. and George, J., Research into Electronic Maps and Automatic Vehicle Location. Eighth International Symposium on Automation in Cartography, AUTO-CARTO 8, Baltimore, Maryland, March 30 - April 2 (1987).

Sedgewick, R., Algorithms, Addison-Wesley Publishing Company, Inc. (1983).

Skomal, E.N., Automatic Vehicle Locating Systems, Litton Educational Publishing, Inc. (1981).

Wiess, S.M. and Kulikowski, C.A., A Practical Guide to Designing Expert Systems, Rowman & Allenheld, Totawa, N.J. (1984).

Zadeh, L.A., Fuzzy sets as a basis for a theory for possibility, Fuzzy sets and systems, New York, North-Holland (1978).

Zavolli, W.B., Honey, S.K. and White, M.S., A land vehicle naviqation system supported by digital map data base: ADPA Artificial Intelligence and Robotics Symposium, November 6-7 (1985).