社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2003－ＩＣＳ－133（28）
2003／9／16

Architecture of Spatial Data Warehouse for Traffic Management

Hiroyuki KAWANO† and Eiji HATO††

† Department of Systems Science, Graduate School of Informatics, Kyoto University, Yoshida Hommachi, Sakyo-ku, Kyoto, 606–8501 Japan
†† Urban Environmental Engineering, Faculty of Engineering, Ehime University, Bunkyo-cho 3, Matsuyama-shi, 790–8577 Japan
E-mail: †kawano@i.kyoto-u.ac.jp, ††hato@en2.ehime-u.ac.jp

Abstract Recently, huge volume of spatial and geographic data are stored into the database systems by using GIS technologies and various location services. Based on the long term observation of person trip data, we can derive patterns of person trip data and discover trends of actions by executing spatial mining queries effectively. In order to improve the quality of traffic management, traffic planning, marketing and so on, we outline the architecture of spatial data warehouses for traffic data analysis. Firstly, we discuss some fundamental problems in order to achieve a data warehouse for person trip data analysis. From the view point of traffic engineering, we introduce our proposed route estimation method and advanced spatial queries based on the techniques of temporal spatial indices. Secondly, in order to analyze the characteristics of person trip sequences, we propose the OLAP (On-Line Analytical Processing) oriented spatial temporal data model. Our $\sum$-tree data structure is based on the techniques of data cube and 3DR-tree index. Finally, we evaluate the performance of our proposed architectures and temporal spatial data model by observing actual positioning data, which is collected by location service in Japan.

Key words Traffic Planning, Temporal Spatial Data, Spatial Information Model, Location Service

交通計画のための空間データウェアハウス構成技術

河野　浩之†　　羽藤　英二††

† 京都大学大学院情報学研究科　〒606–8501 京都市左京区吉田本町
†† 愛媛大学工学部環境建設工学科　〒790-8577 松山市文京町３番
E-mail: †kawano@i.kyoto-u.ac.jp, ††hato@en2.ehime-u.ac.jp

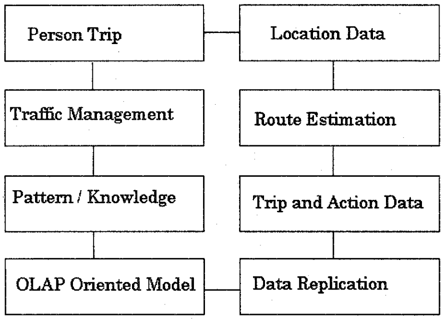
あらまし　PHS,GPS 機能付き携帯端末や各種センサーなどから得られる位置情報を用いた位置情報サービスが普及しており，交通計画に活用可能な位置情報システムも高度化しつつある．そこで，本稿では，データベース関連の最近の技術進展を踏まえながら，交通データウェアハウスに長期間のトリップデータを格納することを提案し，また，効率の高い空間マイニング問合せの実行を行うための OLAP 指向空間情報モデルを導入する．まず，空間インデックス技術に基づいて，収集されたトリップデータに対する最も基本的な処理である経路推定手法について述べる．また，実時間解析を想定したデータ構造として，時空間インデックス技術を発展させた $\sum$-木構造による解析手法について述べる．最後に，都市空間における交通行動調査を行った実データを用いて，交通計画に必要なデータウェアハウスとデータマイニングの可能性についての評価を行う．

キーワード　交通計画，時空間データ，空間情報モデル，位置情報サービス

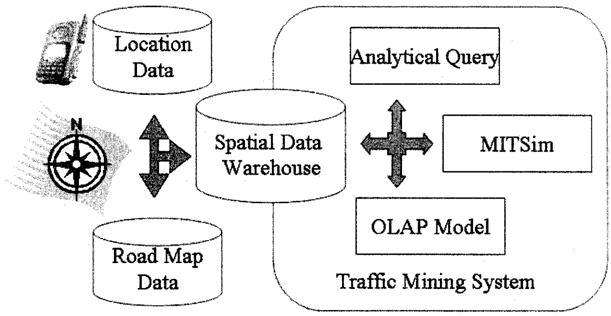
1. Introduction

In recent years, application systems of GIS (Geographic Information System) and spatial databases are becoming popular [13], and we have to operate and analyze the huge volume of spatial temporal data in location service systems based on GPS (Global Positioning System) and PHS (Personal Handy-phone System).

Furthermore, in the research fields of data mining, a lot of spatial data mining algorithms to derive patterns and dis-



(a) Analytical processes for spatial temporal data.



(b) Spatial data warehouse and traffic mining system.

Fig. 1 Processes and Architecture for Person Trip Data Analysis.

cover knowledge in the huge volume of databases are proposed [1], [6], [8], [9]. For example, in order to make clusters effectively, clustering algorithms make full use of the spatial characteristics, such as density, continuity and so on. In our previous researches [3], we also proposed effective clustering algorithms based on the spatial index technologies [11], [12], such as R-Tree, R*-Tree, PR-Quadtree and others.

Therefore, in this paper, we outline an architecture of spatial data warehouse for traffic management/control, traffic planning and marketing, in order to analyze the characteristics of traffic flows based on location services.

In Section 2, we point out several fundamental problems [3], which must be solved before constructing the spatial data warehouses. In Section 3, in order to derive trip routes from actual discrete positioning data effectively, we introduce the route estimation algorithm [4]. In Section 4, in order to execute large-scale traffic simulation, we apply data replication algorithm and estimate more accurate traffic parameters. We also discuss the performance of our proposed method by using actual positioning data by PHS location service in Sapporo city. In Section 5, from the view points of spatial data mining, we discuss typical OLAP queries for traffic mining, and we propose spatial temporal data structures. Finally, we conclude in Section 6.

2. Spatial Data warehouse for Traffic Management

In order to construct data warehouse for traffic management systems, firstly we have to integrate the technologies of GIS, spatial database and location services. In this section, we point out several problems of geographical information systems and location services.

2.1 Architecture of spatial data warehouse for traffic management

We introduce the analytical processes for person trip data in Fig.1(a) and we describe the data warehouses and related processing modules for person trip data analysis in Fig.1(b). Firstly, by using location services, we observe data of "*person trip*" and translate from raw data to "*location data*". We estimate a trip route by using discrete location data and map data for a specific person trip. The sequence of "*trip and action data*" is derived from data of "*route estimation*" and action attribute values. If we have a sufficient volume of person trip data, it may be possible to analyze "*OLAP oriented model*" directly, and we can derive "*pattern and knowledge*" for "*traffic management*" from data. However, due to the problems of privacy and location observing costs, it is very difficult to capture sufficient numbers of person trip data in a target area. Therefore, we apply "*data replication*" algorithm to original observing trip data, then we execute simulation tools (MITSim) by using cloning data of actual person trip.

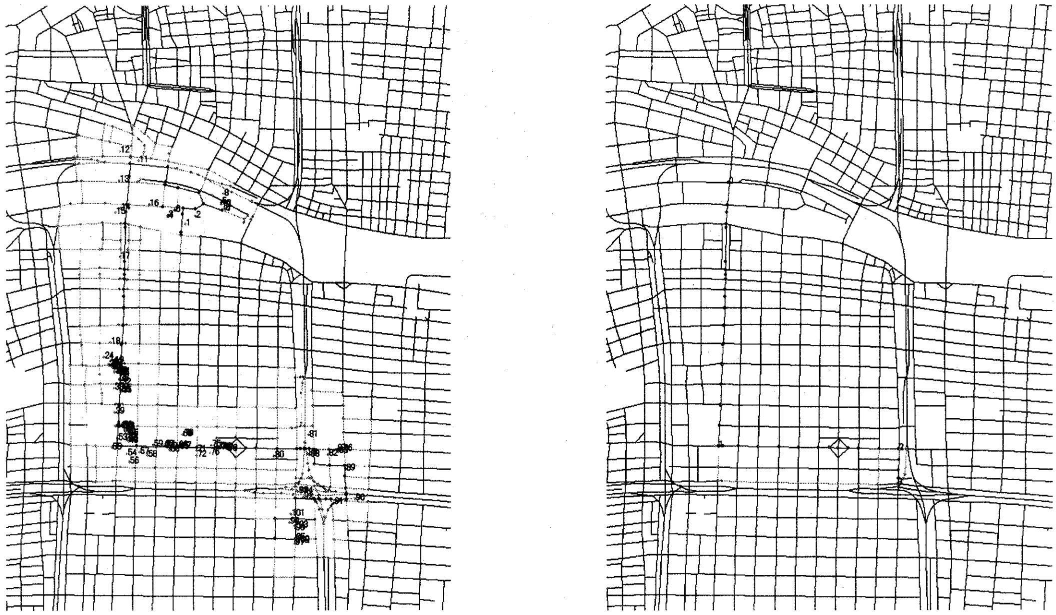
2.2 Basic problems of GIS and location service

Integration of different types of GIS:

There are so many geographic information systems, which are composed of spatial databases and advanced analytical tools. However, it is not so easy to integrate different types of information systems in order to develop the spatial data warehouse for traffic management. Because, we have to integrate various spatial data with different formats with different accuracy.

Clearing warehouses of geographic data:

Clearing warehouses and common spatial data formats, which are sometimes described in XML, are becoming very useful in order to exchange different types of spatial and non-



(a) Positioning sequences and exact route.

(b) Primary estimated route and alternative route.

Fig. 2 Comparison between exact route and estimated route.

spatial attribute values. However, at present, it is very hard to integrate schemata, attributes and many other characteristic values. For example, in order to display the position on a map, we have to handle several different spatial coordinates such as WGS-84, ITRF(International Terrestrial Reference Frame) and others.

Advanced trip monitoring systems:

We evaluated the measurement errors by using mobile GPS terminals [2]. But it is not sufficient to determine the location by using PHS type location services. The major error factor of the PHS service seems to be caused by the reflection of buildings and constructions. Of course, the error is becoming smaller by GPS and pseudolites gradually. Therefore, we have to pay attention to advanced location services[Note1].

3. Estimation of Person Trip Route

In this section, we focus on the estimation of person trip route, and we examine the accuracy and validity of our proposed algorithm by using actual PHS location service.

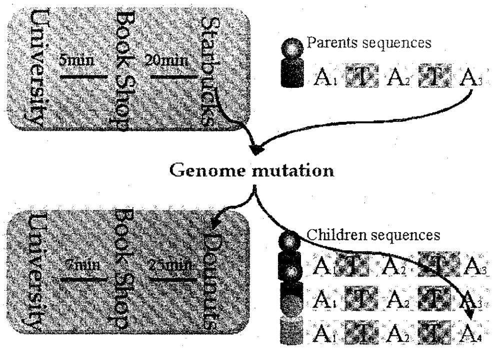
After integrating the numerical map, various geographical and spatial attributes, such as nodes of crossroads, road arcs, directions and so on, spatial objects are stored into the spatial data warehouse. In this experiment, we utilized a numerical map of a restricted area. We also analyzed the sequences of positioning data during 30 minutes with 15 seconds interval of location service.

Fig.2 (a) and (b) show the comparison with the actual sequences of PHS positioning data and the estimated person trip routes that are derived by our Quad-tree based estimation algorithm [4].

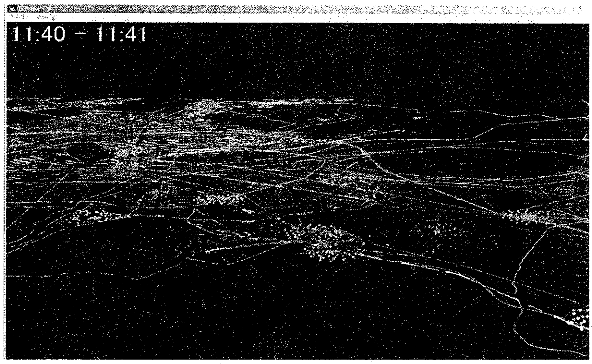
In Fig.2 (a), the numbers show the sequences of positioning data provided by PHS location service, pale -blue- lines show the roads that should be searched, and the red -or black- line with points shows the exact person trip route. In Fig.2 (b), a line with points shows the primary person trip route that is estimated by out algorithm, and several pale lines mean alternative rational person trip route. In this case, the primary route is entirely consistent with the exact route. But the both of estimated routes are also rational, which satisfy traffic legal restrictions. The alternative routes may be rational within the error range of PHS location service. In several experiments, almost of all routes could be specified fast and correctly.

By using our proposed system, it is possible to collect sufficient volume of person trip routes into the spatial data warehouse in order to analyze the traffic congestion and patterns for statistical or mining objectives.

Note1: One of important URLs is http://www.fcc.gov/911/.



(a) Replication of spatial information data.



(b) Visualization of MITSim results.

Fig. 3 OLAP Oriented Spatial Data Structure.

4. Replication of Person Trip for Large-scale Simulation

There are several major analytic methods for traffic congestion, such as stochastic user equilibrium assignment model, numerical computing by simulation model and others. However, it is so hard to consider interfering constraints between trip parameters. Generally speaking, a sequence of person trips and actions strongly dependents each others, but almost of all analytical models don't care those interference of sequential actions.

On the other hand, by using advanced location service, it is to easy to capture the long-term sequence of detail trips and actions. Therefore, we propose our continuous person trip model with interfering actions, an array of observable trip-action is given by sequences of $OT_i = [A_1TA_2TA_3TA_4\cdots]$ in Fig.3 (a). A person trip is presented by a sequence of actions A_i and trip time T.

As we mentioned in previous section, due to the problems of privacy and observing cost, it is very difficult to collect sufficient numbers of person trip data. In order to execute MITSim as a large-scale simulation in Fig.3 (b), we need much more volume of observing data or their replications as initial settings. Therefore, with preserving the order of actions, we replicate and clone the array of trip-action data with different distribution of moving speed and staying time. Easily speaking, the sequence of $OT_i = [A_1TA_2TA_3TA_4]$ is preserved, we produce some mutants of a trip-action sequence presented in Fig.3 (a).

In our experiment, we collected the sequences of 99 person trip data, who had trip-actions from Sapporo city to Sapporo dome on November 24 in 2001, by observing PHS mobile terminals. We visualized sequences of person trips with 5 minutes interval time in Fig.4, we recognized that moving objects were concentrated into specific area, "*Sapporo dome*". Therefore, the replicants and mutants of these sequences are also concentrated in the specific area, when we execute cloning process of trip-action data.

Next, we executed MITSim as large-scale simulation presented in Fig.5. Based on the experimental results, it is possible to estimate the peak value of congesting situation correctly. There is a little bit delay of trips between the actual observation and our simulation results.

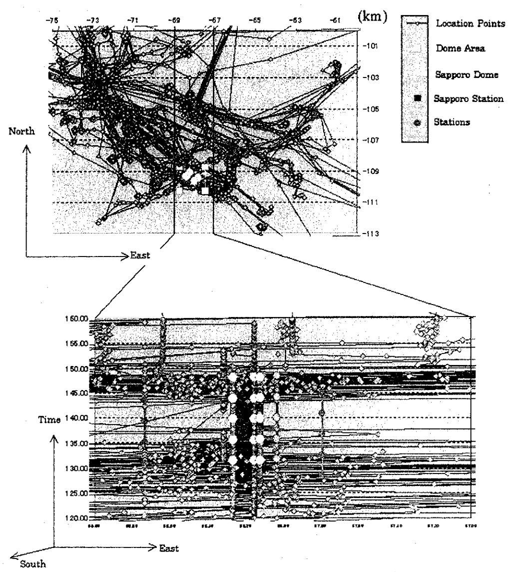
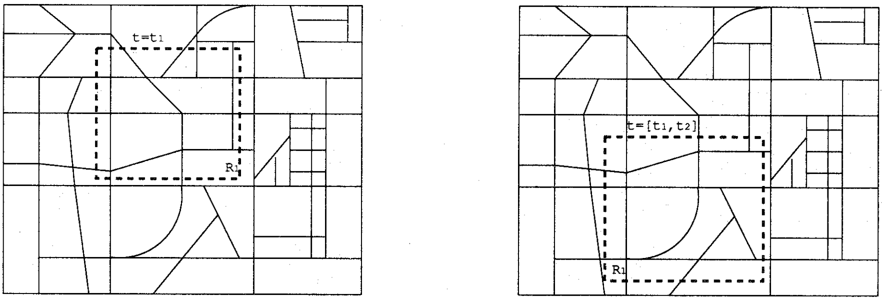


Fig. 4 Visualization of Person Trip Data.

Next, it is important to analyze and discover the primary



(a) Timeslice query.

(b) Window query.

Fig. 6 Spatial Temporal Queries for Traffic Analysis.

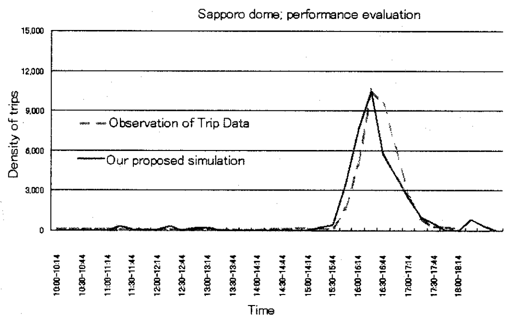


Fig. 5 Comparison of observation and simulation results.

factors of this congesting situation in order to have traffic management and control.

5. Analytical Queries in Spatial Data Warehouse

In this section, we focus on the problems of OLAP (On-Line Analytical Processing) in a spatial data warehouse for traffic management and control. We discuss important analytical queries from the view point of traffic engineering. We also propose OLAP oriented spatial data structure in order to derive characteristics of moving objects in the data warehouse effectively.

5.1 OLAP queries for traffic management

Firstly, we have to consider temporal spatial index such as TPR-tree (Time Parameterized R-tree) [10] in order to handle moving objects dynamically [7]. TPR-tree is an extension of R-tree index, and moving objects are stored in nodes of TPR-tree index by using the time function.

After we store a huge number of trip-action sequences into a spatial data warehouse, we need to execute temporal spatial queries [10] in order to discover characteristics of traffic flows from the view point of traffic management and control. Here, we use definitions of time series, t_1, $t_2 (t_1 < t_2)$, and regions, R_1, R_2, and the following temporal spatial queries are important for our analysis.

(1) **Timeslice query** $Q_{ts} = (R_1, t_1)$: At time point t, objects are searched for in a region R_1 in Fig.6 (a).

(ex.) Based on the results of a query, we can calculate typical traffic flow parameters, such as traffic density, average traffic velocity and others.

(2) **Window query** $Q_{win} = (R_1, t_1, t_2)$: Fig.6 (b) shows that moving objects are searched for in the region R_1 from t_1 to t_2.

(ex.) By using the results of window queries, we can calculate time average velocity which has rather stable property in traffic analysis.

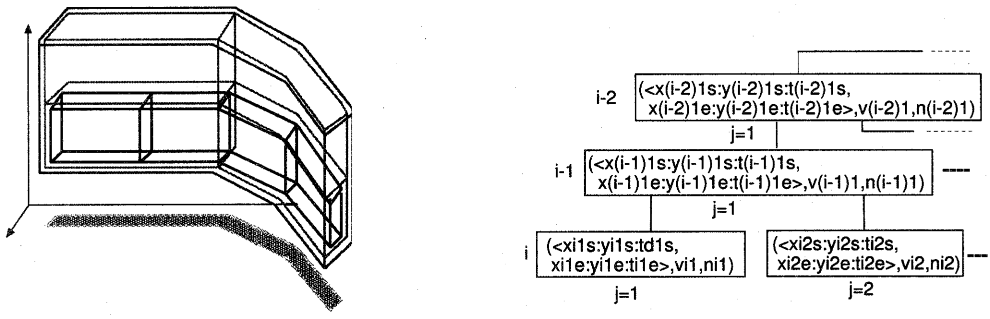
5.2 $\sum$-tree for person trip data analysis

In order to execute analytical queries and mining processes, we proposed our spatial temporal data structure, which is based on the technologies of spatial temporal indices and data cube. Our proposed $\sum$-tree data structure in Fig.7 has a hierarchical tree structure with having total number of objects and sum of objects' speed stored in lower nodes [5].

For instance, we consider spatial temporal data $(x_1, y_1, t_1), (x_2, y_2, t_2), \cdots, (x_n, y_n, t_n)$ with total number of objects n. This hierarchy is based on the spatial constraints with moving speed and direction of objects. Nodes $L_1, L_2, \cdots, L_N$ are constructed hierarchically by using spatial constraints of objects, such as similar vectors with moving speed and direction, and objects are stored in same segments of roads.

$$\begin{cases} (x_1, y_1, t_1), (x_2, y_2, t_2), \cdots, (x_i, y_i, t_i) \subset L_1 \\ (x_{i+1}, y_{i+1}, t_{i+1}), \cdots, (x_j, y_j, t_j) \subset L_2 \\ \qquad\qquad\qquad\qquad \vdots \\ (x_{k+1}, y_{k+1}, t_{k+1}), \cdots, (x_n, y_n, t_n) \subset L_N \end{cases}$$

Furthermore, considering additional speed attribute values $(x_1, y_1, t_1, v_1), (x_2, y_2, t_2, v_2), \cdots, (x_n, y_n, t_n, v_n)$ in a leaf



(a) Data partitions for trip data analysis.

(b) Objects stored in $\sum$-tree.

Fig. 7 OLAP Orinted Spatial Data Structure.

node, sum of objects' speed $V_{L_1}, V_{L_2}, \cdots, V_{L_N}$ are also stored in upper nodes.

$$V_{L_1} = \sum_{l=1}^{i} v_l,\ V_{L_2} = \sum_{l=i+1}^{j} v_l,\ \cdots, V_{L_N} = \sum_{l=k+1}^{n} v_l$$

For example, in our proposed data structure of Fig.7 (b), we define the specific spatial node ($< x_{ijs} : y_{ijs} : t_{ijs}, x_{ije} : y_{ije} : t_{ije} >$) based on two different tips of nodes $(x_{ijs}, y_{ijs}, t_{ijs})$ and $(x_{ije}, y_{ije}, t_{ije})$. We also store area of nodes $(x, y, t) \subset L_u$, sum of traffic parameters, such as total speed of objects V_{L_u} and total number of objects N_{L_u}. We can store those values into nodes recursively. We name this spatial temporal data structure as $\sum$-tree, in the specific nodes including moving objects, it is possible to calculate summing-up and average values by using this structure with small computing cost effectively [5].

6. Conclusion

In this paper, we outlined the framework of our proposed spatial data warehouse for traffic management/control and discussed typical queries for traffic data analysis and mining. At present, in order to analyze the traffic flow and discover complex patterns from huge volume of person trip-action sequences, we need to execute our proposed cloning process in our architecture. In near future, if we can collect all actual positioning data in a region, we may omit the replication and simulation modules in our proposed system.

Acknowledgment

A part of this paper is dependent on the research supported by Grant-in Aid for Scientific Research of the Ministry of Education, Science, Sports and Culture. And a fundamental part of this work is supported by the grant of Mazda Foundation.

References

[1] Ester, M., Frommelt, A., Kriegel, H.-P., and Sander, J., "Algorithms for Characterization and Trend Detection in Spatial Databases", Proc. of the fourth ACM SIGKDD International Conference (KDD-98), pp.44-50, 1998.

[2] Hanshin Expressway Public Corporation, "Technical Trends of Mobile Location Systems", Technical Report of Mobile Location Service, February, 2000. (In Japanese)

[3] Ito, Y., Kawano, H., Hasegawa, T., "Index based Clustering Discovery query for Spatial Data Mining," Proc. of the 10th Annual Conference of JSAI, pp.231–234, 1996. (In Japanese).

[4] Kawano, H., "Architecture of Trip Database Systems: Spatial Index and Route Estimation Algorithm," XIV International Conf. of Systems Science, Vol. III, pp.110-117, Poland, 2001.

[5] Kawano, H., "On-line Analytical Processing for person trip database," The sixteenth triennial conference of the International Federation of Operational Research Societies, p.118, 2002.

[6] Koperski, K. and Han, J., "Discovery of Spatial Association Rules in Geographic Information Databases," Proc. 4th International Symposium SSD '95, pp. 275–289, 1995.

[7] Minami, T., Tanabe, J. and Kawano, H., "Management of Moving Objects in Spatial Database: Architecture and Performance," Technical Report of IPSJ, Vol.2001, No.70, DBS-125, pp.225–232 (2001). (in Japanese)

[8] Ng, R.T. and Han, J., "Efficient and Effective Clustering Methods for Spatial Data Mining," Proc. 20th VLDB, pp. 144–155, 1994.

[9] Rogers, S., Langley, P., and Wilson, C., "Mining GPS Data to Augment Road Models," Proc. of the fifth ACM SIGKDD International Conference (KDD-99), pp.104–113, 1999.

[10] Saltenis, S., Jensen, C. S., Leutenegger, S. T., and Mario A. Lopez, "Indexing the Positions of Continuously Moving Objects," Proc. of the 2000 ACM SIGMOD International Conference on Management of Data, pp.331-342, USA, 2000.

[11] Samet, H., "Spatial Data structures," Modern Database Systems, (W. Kim, ed.), ACM Press, New York, pp. 361–385, 1995.

[12] Samet, H., "The Design and Analysis of Spatial Data structures," Addison-Wesley, Reading, Mass. New York, 1995.

[13] Shekhar, S. and Chawla, S., "Spatial Databases, A tour," Prentice Hall, 2003.