MODELING OF CHANGE-ONLY INFORMATION AND UPDATING OF SPATIAL DATA

YING shen

School of Resource and Environment Science, Wuhan University. Wuhan .430079 .China - shy@whu.edu.cn

KEY WORDS: Geographical Iinformation Science, Spatiotemporal Modelling, Geo-information Management, Database Updating, Change Detection

ABSTRACT:

We have constructed multi-scales foundational geographical database in china, and have released diverse spatial data and derived products. But how to keep the currency and continual updating of spatial data becomes the key problem. Now we almost update geographic information in batch volume mode. Data vender provides updated data to client users and the users update their own database according to the updated data. But this approach shields the concrete contents of geographic feature changes, as well as the difference of semantics, structures of geographic information between the vendors and clients. So batch updates method is time-consuming and troublesome with low efficiency. Incremental updates based on change-only information provide a novel approach for geographic data updating. Incremental updates only record the change-only information and characteristics in master database between different version datasets, and deliver the delta information to client users and integrate them into client database. The paper develops the technique workflow of extraction and distribution of change-only information based on different version databases. We take road features as an example to analyze the change types of realistic entities and discuss the object match method in the process of extraction of change-only information. Also we give the definition of change-only information and the idea about distribution and issue of change-only information to update user's database.

1. INTRODUCTION

With the rapid development of economy and society, land covers and landscapes take place huge changes. Geographic information not only has spatial location characteristics, but also the temporal characteristics. On one hand geographic information records the entities of earth surface, on the other hand it reflects the currency and dynamics of geographic features, which make the updating and maintenance of spatial database to be arduous tasks. Because of the huge investments and complex specifications of process and updating of geographic data, we set up many professional companies and official sectors to manage and deal with spatial data, such as china state bureau of surveying and mapping (SBSM), NIMA, USGS, TeleAtlas. Up to date, china has taken shape of foundational geographic information and databases, has established multi-scale, diverse products of spatial data. But how to keep up with the currency of foundational spatial databases and continuous updating have become the hotspots and difficult problems (Chen Jun, et al. 2003).

The future geographic information services will be “on demand and timely updating”, and they include change monitoring, change detection, database updating and products renewing. (Chen Jun, et al. 2007). In the environment of geographic information share, the data vendors with master database provide spatial reference data to other organizations and users, while the client users process and adjust geographic data to integrate with their own database according to their requirements. In order to keep the currency of spatial data, the client users must purchase the data in time from the vendors if they can provide. This has to meet with the problem of discover the change in master database and update the client database. The common method is batch updating, which has drawbacks of long period, high cost, and actually there are only small data have changed in the holistic area. Also the client users must face the risk of inconsistence and lost of their own geographic data with batch updating. Change-only updating gives a novel idea to update geographic database. It detects the difference between different version data in the master database and only records the changes information of geographic features and the corresponding characteristics, and delivers the change-only information to client users with delta files to fusion with client database. The change-only updating has the advantages of high efficiency, high speed and only takes place on limit geographic features with small volume data, which put change-only updating to the front hotspot.

Now china has built foundational geographic databases with different versions. We develop the overall technique framework of detection of change-only information and change-only updating based on versions database to detect, extract, manage and publish the change-only information efficiently, which is illustrated in section 2. Geographic data represent the realistic geographic entities, so we will describe the changes in reality and digital environment in section 3. The paper takes the road feature as example, discusses the change taxonomy based on version database. In section 4 we develop the match approach to extract the change-only information. Also we give the ideas of modelling, transmission, updating integration of change-only information.

2. FRAMEWORK OF CHANGE-ONLY INFORMATION AND UPDATING

Here we define the difference between the new version database and old version database as change-only information. Generally we can produce the change-only information by subtracting the old database from new database and the mutual comparison. Detection and representation of change-only information is the key of change-only updating. To get change-only information we must specify the baseline, here we use some certain old version map/database as baseline map. So change-only information is different according to different baseline map. In principle we can designate many baseline maps according to different applications to reduce data volume. We can use change-only information based on baseline map to reduce the volume of updating data. Change-only updating model is based on incremental information description according to baseline map and user data status, and the updating must be carried out to add change-only information on client data. In result the updating will result in modification of customer database to be consistent to current spatial data.

Change-only updating becomes the trend of spatial data updating because of its model flexibility and guarantee of geographic data currency. Change-only updating of spatial database includes three components, that's detection and extraction of change-only information based on version databases, management and distribution of change-only information, updating of client database. Collection of change-only information is the foundation of change-only updating. According to the given national foundational information system, the paper finds the changes and change-only information through the overlay analysis and contrast of different snapshot with different version databases. We name the old and new version database respond to time stamp.

The main framework of change-only information and updating can be delivered by figure 1. We first normalize the version databases to determine the candidate geographic datasets to match the corresponding objects. Then we analyze the relationships and identify the change taxonomy among objects between old version database and new version database through geometric, thematic and semantic analyses. With the definition and data structure of change-only information we then can detect and extract the changes of geographic data, and store and manage the change-only information. Last we can publish the change-only information as standard exchange file to update and integrate with client databases.

There are many key techniques and difficulties in the change-only updating. First we should build rapid index and query mechanism to locate and match spatial objects efficiently in different version databases with the very large volume geographic data. Fortunately there are many spatial index and database techniques that we can utilize, such as R-tree, Oracle software. The second problem is to develop efficient method to extract change-only information. We use object match method to determine the difference between old objects and new objects through geometric, social and spatial properties. We consider that there are quantitative changes and qualitative changes with the changes of geographic features. It's not enough to distinguish them only by the geometrics of spatial objects. We should take the semantics and neighbourhood of spatial objects to the changes type. The representation and modelling of change-only information is also the key problem, which should resolve the representation difference of objects in old database and new database to support extraction and storage of change-only information. Importantly, we must insure that the client users can identify and parse the change-only information to integrate it with their own database.

3. DESCRIPTION OF CHANGE TYPES

Before the extraction of change-only information, the corresponding change relationships of realistic entities before and after change should be made clear. Spatial-temporal model based on change-only information focuses on the changes or events of spatial objects, and the corresponding relationships of spatial objects should be matched and analyzed. These corresponding relationships is analyzed based the representation of spatial objects based on different version. In the users' view, spatial data updating is the conceptual description of changes with geographic entities. In other words, the users pay more attentions to changes of geographic semantics. Claramnut et al. (1996) gave three types changes: basic processes, transformation processes and movements that include eight detail types. This analysis is suitable for single spatial object, but it can not deal with the multi objects and groups in spatial-temporal change. For road element in the paper, we take Claramnut's view as reference to analyze the spatiotemporal road features, and take multi-objects into consideration, analyze the change states of realistic entities, divide the changes into seven types: appearance, disappearance, splitting, connection, reshape, semantics change and modification, stability. Table 1 illustrates the details changes of spatiotemporal road features with the descriptions and examples.

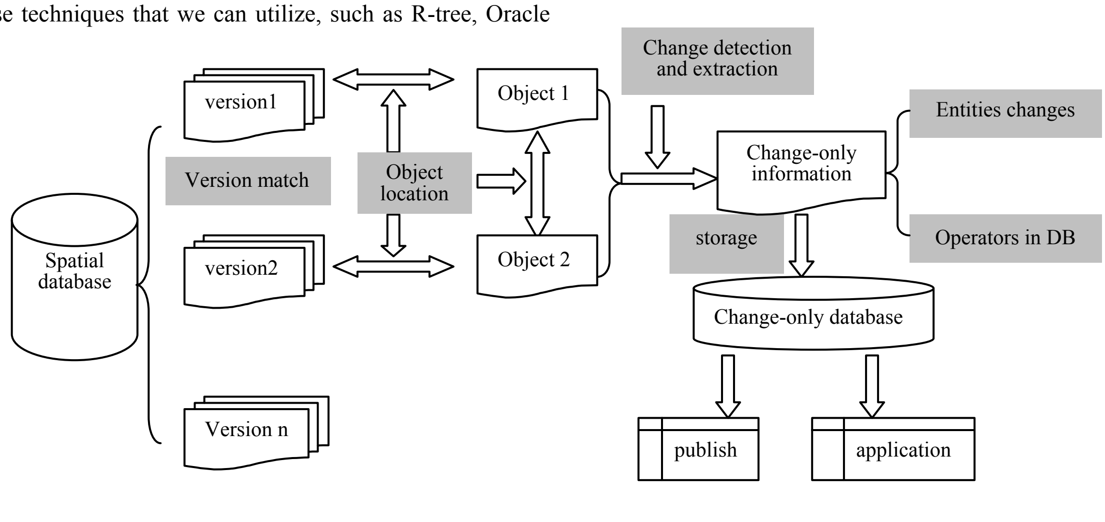

Figure 1. Framework of change-only information and updating

	Change type	description	example	operators
Single object	Geometric reshape	extension		Modify
		shortening		Modify
		reshaping		Modify
	Semantics transformation / thematic change	reclassification	Class 3-> Class 2	Modify
		renaming	Lawson St. -> Luoyu St.	Modify
		widening	20M -> 30M	Modify
		Lane change	4 lanes -> 6lanes	Modify
		Condition change	Road condition: gravel -> pitch	Modify
	appearance	New building		Add
	disappearance	abandoning		Delete
Multiple objects	connection	connection		Delete, modify
	split	split		Delete, modify

Table 1. Change types analysis

4. OBJECT MATCH PROCESS

During the change-only updating, detection and extraction of change-only information is the fundamental task, and main approach is based on object match. We can compare spatial objects in semantics, geometric, thematic and topological aspects to determine the difference. Badard (1999) provided a down-to-up method to match the objects, in which geometric node match is processed, then the geometric object and last the semantics. But in this method the candidate sets have much more objects which lead to take long time to detect the change information, so it's not suitable for large volume spatial database. Since we have built national foundational geographic database according to certain specifications and standards, the pattern and format have many similarities. Here we develop the up-to-down method.

At the beginning of object match, we must first identify the database pattern, and select the corresponding version databases and feature semantics to eliminate other irrespective objects to confirm the candidate objects in the right dataset. Version match depends on the metadata of geographic database which include spatial and temporal aspects. Spatial metadata describes the geographic range, coordinate reference system, sheet numbers of topographic maps, and son on. In china, topographic maps are produced and updated in framing map, so spatial metadata match can ensure that the datasets and objects have the consistent scale, sheet number, and layers. Through spatial metadata match we can make the candidate datasets having the comparability. Temporal match refers to the candidate datasets must have time span to conduce to the difference among the temporal datasets.

The candidate objects must have the same or similar feature class before comparison. We call it semantics match in feature level. The two candidate datasets must have the same thematic content during the match, such as road, resident, hydrology. Obviously, the object in one feature can't change to another feature, for example, the road become water. But road elements with class 2 may upgrade its class to 1 because they all belong to the same feature class and reclassify their grade state. These semantics transform can happen on both single object and multi-objects. Figure 2 shows the top-to-down match process.

The common method for object match use geometric match by shape contrasts and topological relationships. If the candidate objects overlap in geometric shape, we can make the further comparison to get the details of change-only information. Geometric match can be processed independent of spatial database. For line object, there many approach and algorithms to deal with these problem, such buffer analysis, hausdorff distance (Abbas,1994), probabilistic statistics and area occupant (Vauglin and Bel Hadj Ali, 1998; TONG Xiaohua et al., 2007) (Volker Walter and Dieter Fritsch,1999). The author (Ying Shen, et al.,2006) presented the matching method of road elements in navigation database based on GDF.

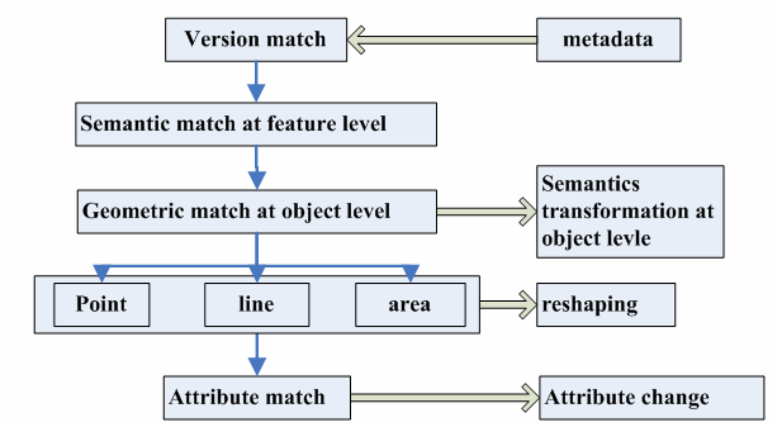

Figure 2. Up-to-down object match

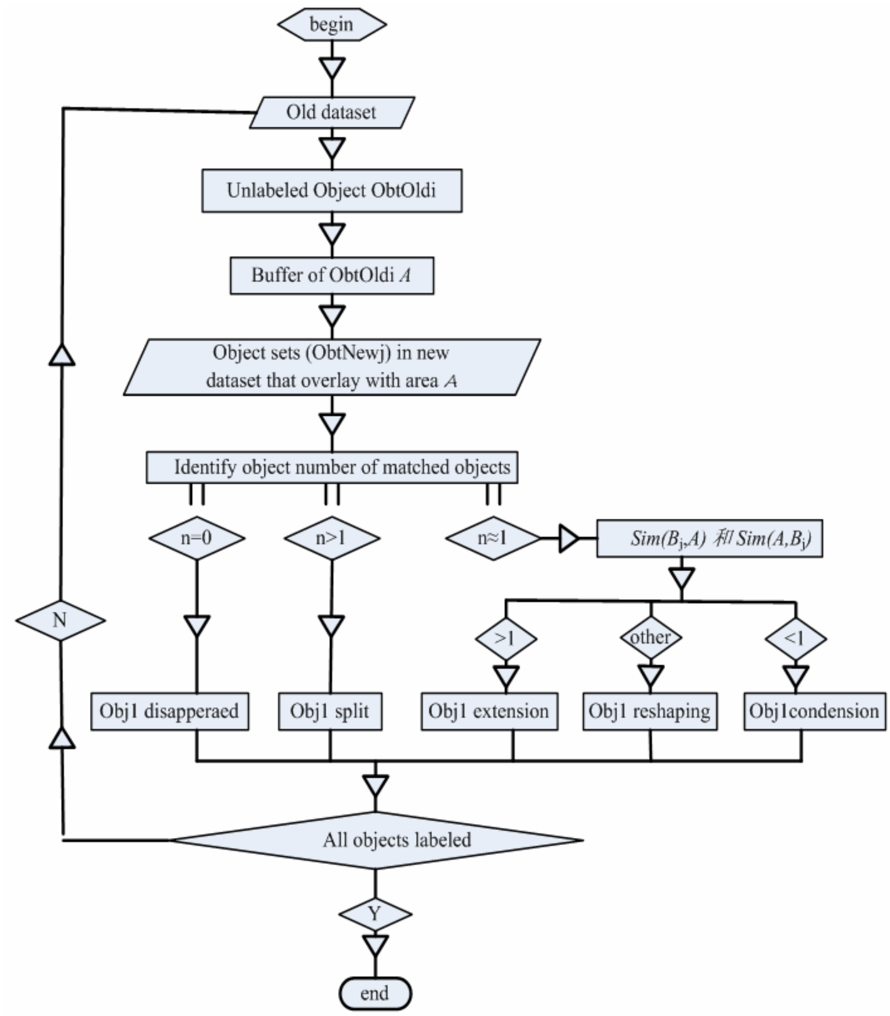

Figure 3. Work flow of change detection with road objects

In this paper we take linear road object as an example to process match. We first build buffer analysis of two candidate datasets, and calculate the area proportion of the overlay. Figure 3 shows the work flow of change detection by unilateral match from old to new version dataset. We select one object "ObtOldi" from old version datasets to build buffer area *A*, and query the candidate objects ObtNewj with buffer area *Bj* in new version database that overlay with *A*, and then determine the change types according to the cardinal relationship which is described in the following. For 1-n cardinality we can say one road object splits into several objects. If they are 1-1 cardinality, we make further computation of area proportions $sim(B_j, A) = overlay(B_j, A)/A$ and $sim(A, B_j) = overlay(B_j, A)/B_j$ to identify the state of extension, shortening and reshaping. Having gotten the candidate matched objects, we can deal with the match at attribute level to extract change details of geometrics and attributes.

There are many cardinal relationships of the matched objects between old version dataset and new version dataset: 1-1, 1-N, N-1,N-M. To get these cardinalities we must make bidirectional match between old and new version datasets. The match from old version dataset to new version dataset bring 1-N cardinal relationship, and vice versa. Multiple object match can avoid the inconsistency efficiently during the updating, which can describe the holistic change-only information. According to the cardinal relationships, the corresponding change events can be described as following.

- 0-to-1: the unmatched objects in new version dataset is described by the new created objects;
- 1-to-0: the unmatched objects in old version dataset is described by the deleted objects;
- 1-to-1: one object in old dataset matched by one in new dataset is described by 1-1 cardinal relationship;
- 1-to-N: one object in old dataset matched by many objects in new dataset is described by 1-N cardinal relationship, which means that the object is split into several objects;
- N-to-1: many objects in old dataset matched by one object in new dataset is described by N-1 cardinal relationship, which means that the objects merge into one object;
- M-to-N: M objects in old dataset matched by M objects in new dataset is described by M-N cardinal relationship, which means that the objects change their structure and regroup;
- Other objects are stable objects.

5. DEFINITION OF CHANGE-ONLY INFORMATION AND UPDATING

The change-only information includes not only the changes about geographic information, but also the relationships between master database and client database and operators in database to be implemented in updating. So we represent the change-only information in three aspects. The first aspect is the difference between spatial objects in semantics, graphics and attributes in the details. The changes in fields domain are recorded, such as the reshape of road with different coordinates. The second component is the relationships among spatial objects before and after updating, and we analyze the relationships at object level and represent them with change events. Also we can define some concepts to describe the change of reality features, such land mergence, road split. In the third component we express the operators information and rules in change-only information at database level in order to carry out the implementation of database updates.

Through the change detection we can get the information about geometric, semantic and attribute change about spatial objects. In order to satisfy the users' query about change, the change-only information must include the event description to express the spatial entities' changes. Also change-only information will be used to update client database ultimately, so the operations in database must be considered in change-only information. So we definite the change-only information with five components, identifier, difference of graphics, semantics and attribute, events.

$$\Delta I=\{ID,\ \Delta semantics,\ \Delta graphics,\ \bigcup_{i=1}^{n}\Delta Attribute,\ event\}$$

Because of the particularity of identifier, we list it in change-only information. But actually we don't need the uniform of identifiers in two databases, and it can be deal as one common item of object. "Δsemantics" represents the difference of semantics between two spatial objects, which mainly involve classification information. "Δgraphics" delivers the difference of geometric shape of two objects. Event includes not only the change information of reality entities, like reshaping, abandoning of road, but also the operators with database to update, which are showed in table 1 . Figure 4 illustrates the result of change-only information with graphics between old version dataset and new version dataset.

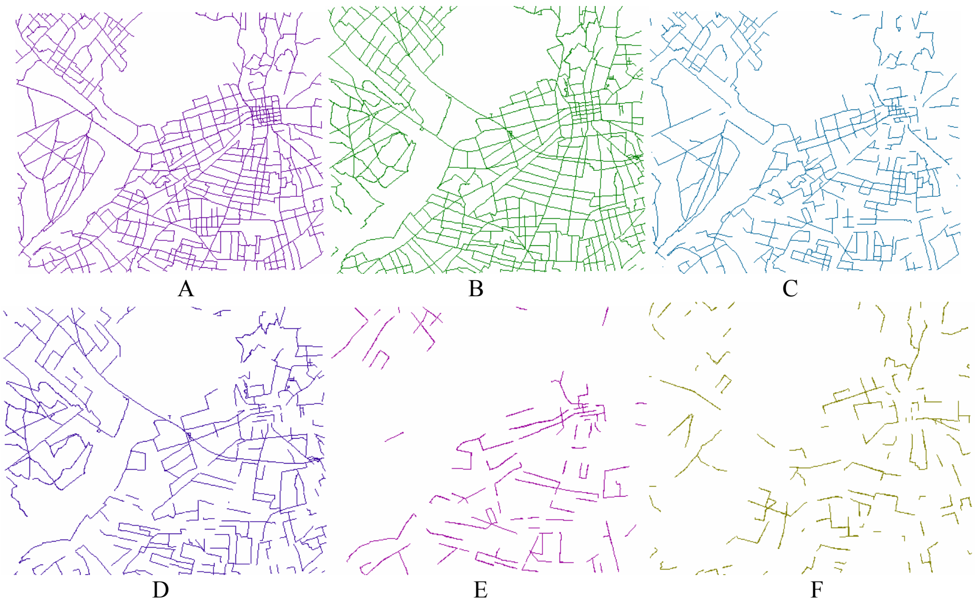

Figure 4. Case study: A)old dataset; B) new dataset; C) deleted road; D)added road; E)connected road; F)reshaped road.

When we perform the updating we must publish and hand out the change-only information to the customers, and there should be some methods to exchange the updating information. In different application fields the physical storage files or data formats are different. We know that there are log files to record the changes of database, and we should create exchange file of change-only information to deliver the updating information. Here we choose the extensible markup language (XML) to express the change-only information because of its standard, openness, readability and flexibility. So the client users can parse the XML files to get the updating information. Figure 5 illustrates the XML schema of change-only exchange files. One road object with ID 301402076 that splits into two objects with ID 301968811 and 301968812 can be described as following:

<Relation RelationName="Split">
<FeatureOldNum>1</FeatureOldNum>
<FeatrueOldID>301402076</FeatrueOldID>
<FeatureNewNum>2</FeatureNewNum>
<FeatrueNewID>301968811</FeatrueNewID>
<FeatrueNewID>301968812</FeatrueNewID>
</Relation>

This complex operation of split can be divided into two creation operations in database to implement the updating. The following shows XML exchange file of new creation of object with id 301972711.

<UpdateOperatrion Operation="Add">
<PolylineFeature FeatID="301968811">
<ID>301972711</ID>
<OffName>Leith St</OffName>
<NumOfPt>2</NumOfPt>
<function>7</function>
<net2class>4</net2class>
<RoadCondit>2</RoadCondit>
<Coordinates>-836798170,430428160 -836793670,430428220</Coordinates>
</PolylineFeature>
</UpdateOperatrion>

Having gotten the exchange file based XML, the client users can update and upgrade their own database with the updating package. The exchange files based on XML don't request the consistence of structure and pattern with database between data centre and clients, and the clients can parse the exchange file to get compatible information to update their system.

The change-only information of spatial data can be produced in data centre, which should provide the updating services. We can use push or pull update strategy to carry out change-only updating through the internet and wireless broadcast. In pull strategy, the clients sent a request to data centre, and data centre need check the version between clients and centre to determine and detect change-only information, and produce the exchange files and transmit them to clients through the internet or wireless communication. But in push methods data provider will automatic sent the change-only information to the clients periodically according to users' subscription. Considering the customer's demand, pull strategy is the prior method because of its reliability and stability. In principle, if we get the change-only information from the same baseline map we can make continue updating with algebra operations of change-only information package to achieve the timely and continue updating as figure 6 shows.. But if the changes are larger than the threshold we must jump to another baseline map to deal with change-only information.

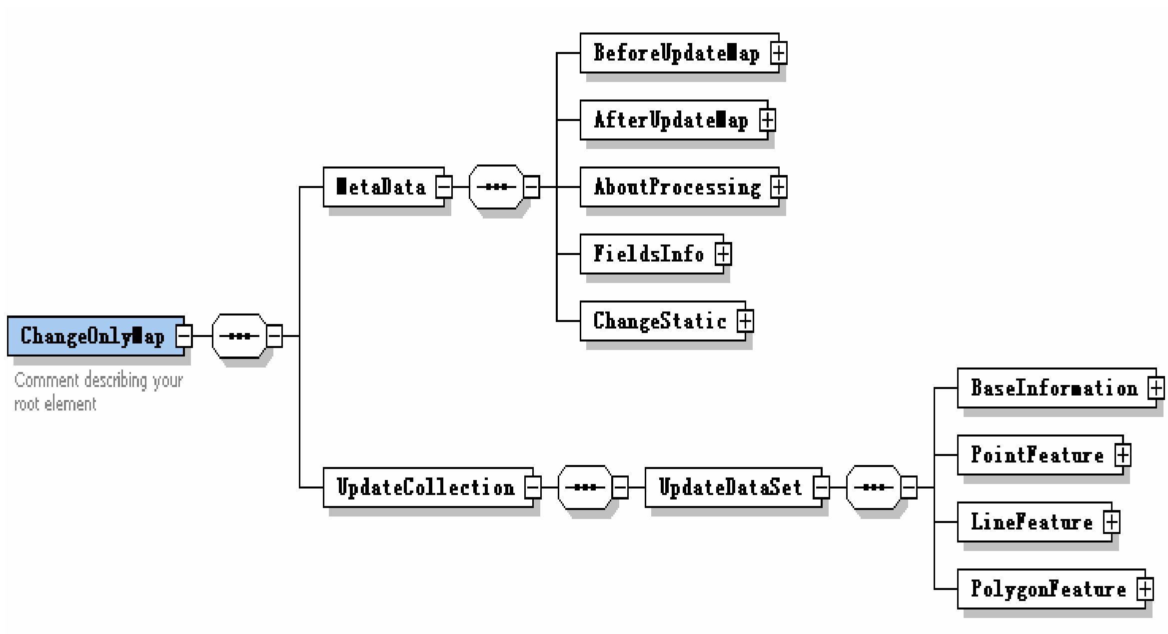

Figure 5. XML schema of exchange file

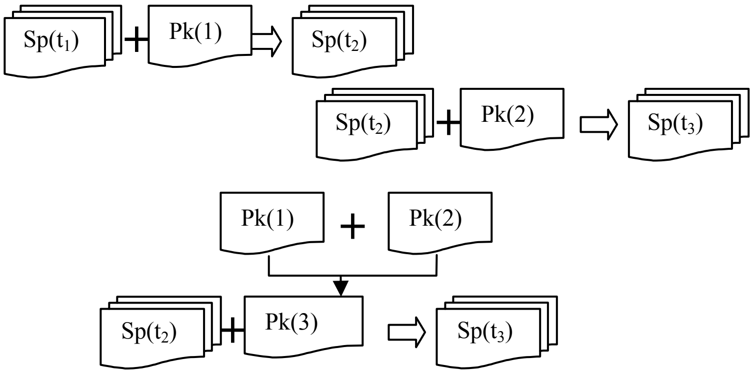

Figure 6. Algebraic operator of change-only exchange file.

6. CONCLUSION

The paper develops the method of change detection based on object match among different version databases, delivers the process of object match and change taxonomy. The change-only information only record the changes happened about geographic objects, and at a limit time span its volume is very small to updating client database quickly and conveniently. Also we don't need the uniform identifier in database system, which is very important for client user because the identifiers in client system are often inconsistent with master database in data centre. And the exchange files based on XML of published change-only information is feasible to implement and integrate.

ACKNOWLEDGEMENT

This research is supported by the National Key Technology R&D Program of China under Grant No. 2006BAB10B00.

REFERENCE

1. Badard T., 1999. On the automatic retrieval of updates in geographic databases based on geographic data matching tools. In *the Proceedings of International Cartographic Association*, Ottawa.

2. Chen Jun, Li Zhilin,Jiang Jie, Zhao Renliang, 2004. Key issues of continuous updating of geospatial Databases. *Geomatics World*. 2(5),pp.1-5

3. Chen Jun, Zhao Ren-liang, Wang Dong-hua.2007. Dynamic updating system for national fundamental GISs: Concepts and Research Agenda, *Geomatics World*, No. 5,pp.4-9

4. Claramunt, C. and Theriault, M., 1996. Towards semantics for modeling spatiotemporal processes within GIS, *7th International Symposium on Spatial Data Handling, SDH'96*, Aug. pp.12-16.

5. Tong Xiaohua, Ddeng Susu, Shi Wenzhong. 2007. A probabilistic theory-based matching method, *Acta Geodaetica et Cartographic Sinica* (in Chinese), 36(2),pp.210-217.

6. Vauglin, F., Bel Hadj Ali, A. 1998. Geometric matching of polygonal surfaces in GIS. In *the proceedings of the ASPRS-RTI, Annual Conference.* Tampa, Floride, USA, pp. 1511-1516.

7. Volker Walter, Dieter Fritsch. 1999. Matching spatial data sets: a statistical approach, *International Journal Geographical Information Science*, 13(5),pp. 445- 473

8. Ying Shen, Li Lin, Liu Xiaofei, Zhao Hu, Li Dan. 2007. Change-only modeling in navigation geo-databases. *ISPRS Workshop on Updating Geo-spatial Databases with Imagery & The 5th ISPRS Workshop on DMGISs*. August 28-29.Urumchi, Xingjiang, China