INTELLIGENCE BASED CADASTRAL DATABASE SELECTION AND VISUALIZATION SYSTEM: CONCEPTS AND PROTOTYPE

Abdullah Hisam Omar and Amran Bachok
Department of Geomatic Engineering
Faculty of Geoinformation Science and Engineering
Universiti Teknologi Malaysia
81310 Skudai , Johor.
Te l : 607-5530945
Fax : 607-5566163
a.hisham@fksg.utm.my

Abstract

The process of extracting land records information especially cadastral data for a large number of cadastral lots was tedious and time consuming. The development of Digital Cadastral Database (DCDB) by Department of Survey and Mapping Malaysia (DSMM) is one of the initiatives to shift to modern cadastral system. Least Squares adjustment approach is one of the main components in realization of Coordinated Cadastral System. Cadastral data need to be adjusted in order to verify the network closure. In order to move to a new adjustment approach, an Intelligence Based Cadastral Database selection and Visualization System should be developed. The system acts as a data provider for adjustment process. The objectives of the study are; i) to develop an intelligence based database integrity and self-correction mechanisms and, ii) to develop cadastral database selection and visualization system. The capabilities of the system are; executing an intelligence database integrity mechanism, generation of zones and overlapping areas also extracting cadastral information from existing DCDB. The system will automatically conducts intelligence cadastral survey concept self- checking procedures, re-compute or rebuild the cadastral records based on mathematical cadastral survey rules, preparation of corrected data input files (DCDB and GPS) for a particular zone and reformatting of data for input into least squares adjustment software. In order to support the development of survey accurate cadastral system, object-oriented programming language approach was adopted. Prototype test bed has been executed and results show that the prototype is having a good reliability and integrity.

1.0   INTRODUCTION

Geospatial data are the product of surveying and mapping which is greatly in demand by users of geospatial data for making analysis on the condition of the earth surface using Land Information System (LIS). Currently there are many problems in getting geospatial data such as; the process of producing geospatial data is tedious, time consuming and incurs high costs. The establishment of CCS is a new initiative undertaken by many advanced countries including Malaysia to address the problems of producing and sharing geospatial data in order to minimize cost.

The rapid development in spatial information infrastructures in many developed countries worldwide is parallel by a move to economic rationalism and micro-economic reform, both of which have had a major impact on the role of spatial data infrastructures. This is to ensure that there are the necessary infrastructures to support users of geospatial data to get accurate information on the true conditions of the surface of the earth. The information referred to be geospatial data, which is data that indicates the true condition of a specific area of the earth surface.

Future advancement in the integration of information in the fields of GPS, GIS and LIS will require the country to convert the existing cadastral system into a "coordinate-based cadastral" system (CCS). This concept allows the an intelligence-based application ability of abundance of survey control stations whose coordinates are given with respect to a geocentric datum, the adoption of a single coordinate projection system for the whole country, and the application of a rigorous network adjustment technique so that a unique coordinates and their respective accuracy information can be defined for every cadastral boundary marks (Majid Kadir et al, 2002 and Ahmad Fauzi, 2002)

CCS offers comprehensive cadastral programs in coping with the IT requirements in the next millennium. This system makes GPS and GIS compatible since coordinates are the basic input/output unit of such systems. It also facilitates the integration of cadastral and map-based information, and the use of rapid data acquisition and storage, processing and management techniques.

Studies on CCS in Malaysia have been conducted since 1996. During the previous studies, all input datasets have been acquired manually using key board entry method. This method is time consuming and labor extensive, and not an appropriate method to be implemented nationwide or when a large cadastral network is concerned.

As, a result an intelligent selection module was developed in order to form and to prepare an input data for a large cadastral network adjustment. The prototype will able to minimize operation at cost, to ease the operation at workload, to reduce labor extensive factor, and to handle and process large amount of data.

2.0 DIGITAL CADASTRAL DATABASE

The most significant change that IT has brought about is the shift from conventional analog data to digital data and consequently the introduction of the concept of digital database, which forms the base component for a Land Information Systems (LIS). This database in turn has been identified as having an indispensable role in national development as it intelligence-based application ds the process of decision making in resource management and planning. Development of the Digital Cadastral Database (DCDB) was initiated in 1986 and commenced with the development of digital cadastral database (DCDB). The Computer Assisted Land Survey System (CALS), took off as a pilot project in the State of Johor. This was subsequently followed by the implementation of another CALS, in the state of Pahang, in 1993. (Abdul Majid Mohamed, 1998). After a successful pilot project in computerizing its cadastral operation in the state of Johor and Pahang, the initial phase of computerization exercise was completed in 1995 with the Nation wide implementation of the Computer Assisted Land Survey (CALS) system. The CALS system not only enables electronic processing of cadastral surveys but also introduced the concept of a Digital Cadastral Data Base (DCDB). This system can be expected to help increase not only the productivity of current systems in operation, but also contribute to a large extent in, as well as benefit from, the CCS implementation.

With the widespread of Geographic Information System (GIS) and the launched of Malaysia Center for Geospatial Data (MaCGDI) in 2004, the database, which forms the primary building block of GIS, is becoming increasingly significant. In order to fulfill the requirements of the GIS users, data capturing which work to populate the DCDB had been given the priority by DSMM. Computerized systems have been established by DSMM in all States of the Peninsular Malaysia. With those systems, new cadastral surveys conducted by DSMM have been electronically processed using the available facilities of the system. The processing eventually led to the output of finalized cadastral survey data in the form of bearing and distance records in the DCDB, besides the generated coordinates. Clearly, this then serves as one of the sources of database populating.

2.1 DCDB Data Structure Overview

Overview of the DCDB data model is based on the existing DCDB model developed by DSMM. A data model is a formal definition of the data required in a GIS. The data model can take one of several forms namely, a structured list and an entity-relationship diagram. The purpose of the data model and the process of specifying the model are to ensure that the data has been identified and described in a completely rigorous and unambiguous fashion. The data model will allow the formal specification for the entities, their attributes and all relationships between the entities for the GIS. There are four steps involved in the database development, which are external assessment, conceptual modeling, logical design, and physical database.

DCDB has utilized the relational data base approach for its development. DCDB consists of three different files or layers which have been presented by three entities such as cadastral lot, boundary mark and boundary stone. All entities are represented by unique spatial object for geographic representation for example polygon for cadastral lot, point for cadastral mark and line is used for boundary line. The design process of DCDB and CCDB consists of five phases as shown in Figure 1 (Abdullah Hisam Omar, 2004).
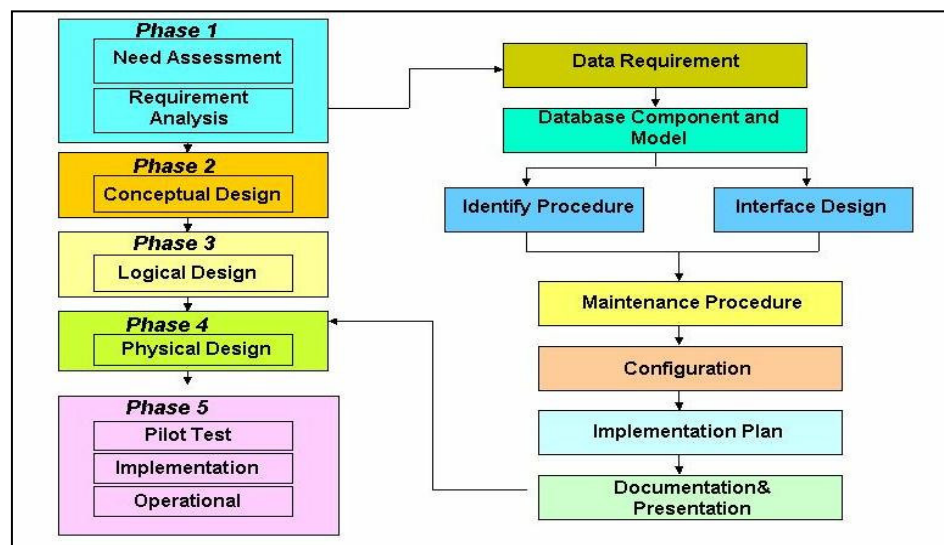
Figure 1. The Design Process of DCDB and CCDB

DCDB conceptual model shown in Table 1 is used to identify data content and to describe data at an abstract or conceptual level.

Spatial relationship (entity-relationship/ E-R) between entities in the DCDB is an explicit recognition of user defined objects, zero or more associated spatial objects and sets of attributes for each defined object. The process of building the E-R diagram involves taking entities from the master data list one at a time and placing each one on the diagram as illustrated in Figure 2 (Abdullah Hisam Omar, 2004). For each new entity, any relationship to any previously entered entity should be entered. Relationships are found by examining the application requirement and determine if the DCDB require a specified operation. The DCDB E-R diagram shown in the Figure 2 will be used to verify the data content. Once verified by the users, the E-R representation can be mapped for a logical database design.

Table 1 : DCDB Master Data List

ENTITY	ATTRIBUTES	SPATIAL OBJECT
Boundary Mark	Pointkey, Apdate, Mark_desc, Serial, Coord_type, R_east, R_north, S_comment, Status, GID	Point
Boundary Line	Apdate, Parcel_key, Bearing, Distance, Units, Class, Line_code, Line_type, Entry_mode, PA, Fnode,Tnode, AdjParcel, Status, GID	Polyline/Line
Cadastral Lot	Negeri, Daerah, Mukim, Seksyen, Lot, Svy_area, Area_unit, Apdate, Status,Lock_ID, GID	Polygon

Logical database design means translation of the conceptual database model into the data model of a specific software system. A logical data model is an abstraction of the objects that we encounter in a particular DBMS application. This abstraction then is converted into database elements. These are the basic elements of the logical data model and their corresponding database elements. Table 2 shows the basic elements of the logical data model while Table 3 shows the logical schema for existing DCDB.
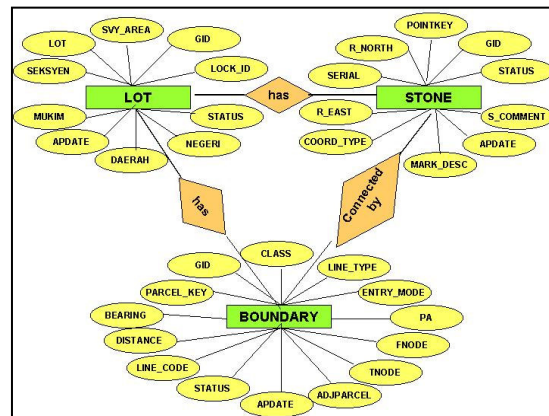

Figure 2. Entity Relationship Model for DCDB

Table 2. Basic Elements of the Logical Data Model

Logical elements	Database elements
Object	Row
Attribute	Column, Field
Class	Table

Database implementation can be carried out after logical scheme was built. The process of physical database design is to represent the logical data model in the schema of the software. Example of the physical DCDB is shown in Figure 3.

Table 3. Logical Scheme for DCDB
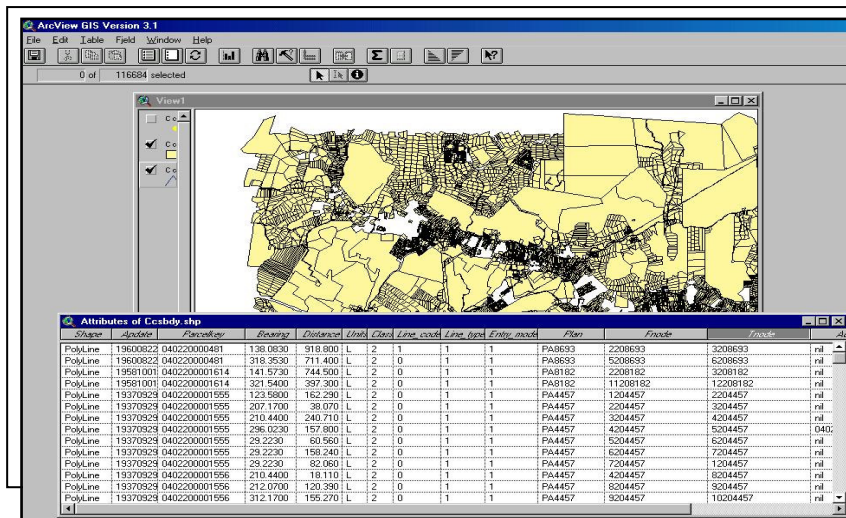
Figure 3. Example of Physical Database Model for DCDB

3.0 CADASTRAL SELECTION AND VISUALIZATION PROTOTYPE DESIGN

The use of computer technology to assist and to solve human problems in various disciplines is increasing rapidly. Computer technology has several advantages over manual method such as the minimization of operation cost, easing the operation at workload, reducing labor extensive factor, as well as its ability to handle and process large amount of data. Computers process data under the control of sets of instructions called computer programs. These computer programs guide the computer through orderly sets of actions specified by the computer programmers. The computer programs that run on a computer system are referred to as software. There are phases involved in the software development as shown in Figure 4.
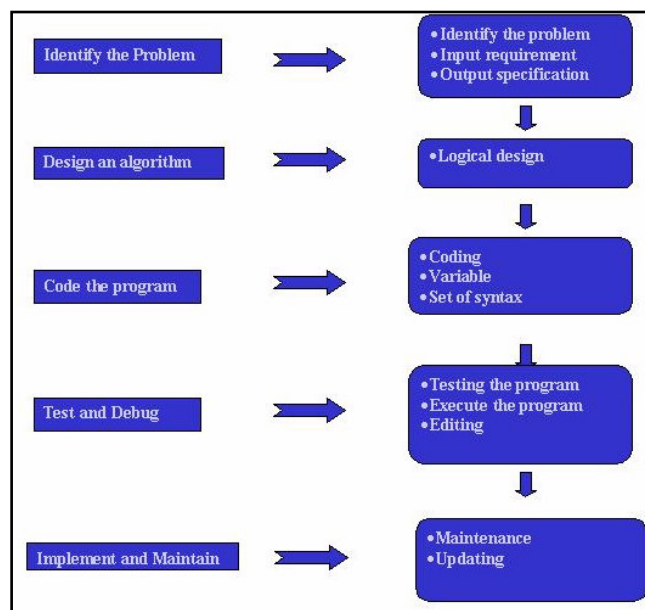
Figure 4. The Software Development Process

3.1 The Need for Computerization

In the previous studies of CCS (Majid et.al, 2002), input data for cadastral network adjustments are bearing and distance for boundary lots, which have been keyed-in manually (Majid Kadir, et.al,1999). This method is not practical to be implemented on the large cadastral network due to tedious task and time consuming. The existence of the State Digital Cadastral Database (SDCDB) that has been developed by DSMM has triggered the idea to develop one system that can be applied leading to automated system for data conversion. The existing DCDB plays an important role towards the success of the CCS study because the state DCDB is the best dataset available for this study. An Intelligence Database Selection and visualization System was designed and developed in Windows-based environment for automation in preparing datasets for large cadastral network adjustment. Furthermore, this system will ease the data selection process and could eventually lead to rapid development of the National Digital Cadastral Database (NDCDB). The prototype application has been developed using the most widely used rapid application development of Visual Basic. In addition, Map Object (ESRI) and Map X (MapInfo) are used to develop an application used to view, analyze, manipulation, and edit spatial data with spatial relation. The integration of object-oriented programming language and developer tools will create the efficient interfacing program to deal with large number of spatial data and its complexity.

3.2.1 Intelligent –Based application

3.2.1 Concept of Prototype Application

Error in GIS or cadastral database is a "fact of life."Care should be taken to keep errors to a minimum at all stages in the construction of a DCDB and it is important to manage them. The key to error management is the ability to track and document the errors. Tedious manual correction procedures must be made to produce an accurate database.

The advantages of intelligence-based application can be exploited in order to solve or to minimize the needs of manual approach. There are many intelligence-based application categories such as Automatic programming, Application and Expert Systems, Robotics, Programming Language and Software. Intelligence-based application can be viewed from variety of perspectives. From computer programming perspective, Intelligence-based application includes the study of symbolic programming, problem solving and search. In other words, generally the intelligence-based application components are knowledge based and method of thinking and problem solving.

The intelligence-based application characteristics are automation and expert knowledge. Object Oriented Programming are a class of languages more recently used for intelligence-based application programming by looking the concept of objects and messages itself, which the object is bundled with data and methods for the manipulating the data. In order to support the development of survey accurate cadastral system, An Automatic Programming and Programming Language approaches will be adopted.

An intelligence-based application database selection system will conduct several outliers integrity checking, rebuild cadastral spatial topology (cadastral lot) and make self-correction procedures based on cadastral survey concepts and mathematical model respective to the cadastral lots selected. This is to ensure that all cadastral lots are kept in a closed polygon and provide accurate and "clean" cadastral information. This scenario is critical in the adjustment process. Cadastral information for a large number of land records can be extracted more efficient digitally than by a manual approach.

The success of the development of National Digital Cadastral Database is depends on the cleanliness of the DCDB and computerization method such as intelligence-based cadastral selection application system. This computerization system is essential and important towards the implementation of computerized Coordinated Cadastral System in Malaysia.

## 3.3   Prototype Design

3.3.1   Data Input

Two databases were used during the application development phases, namely; i) DCDB and ii) Cadastral Control Database (CCDB). Table 4 shows the required data files for this study.

Table 4. Data Input for Application Testing

Database	File	Remarks
DCDB	Boundary Line Connection Line	State DCDB
CCDB	Cadastral Control	GPS Observation

3.3.2   Modeling and Designing Interfacing Program

A proper system design concept is needed in order to achieve the objectives. The conceptual model and unified modeling language are developed to guide the development and the implementation of the system. The interface design involved several tasks, as following; i) the overall amount of functionality including cadastral survey algorithms, filtering algorithm, AI mechanism, navigation, configuration, analysis option; ii) tool appearance will involve types of the design elements such as map images, graphical icons, spatial editor; and iii) arrangement of interface elements means the arrangement of the focal point element (maps) (Figure 5).
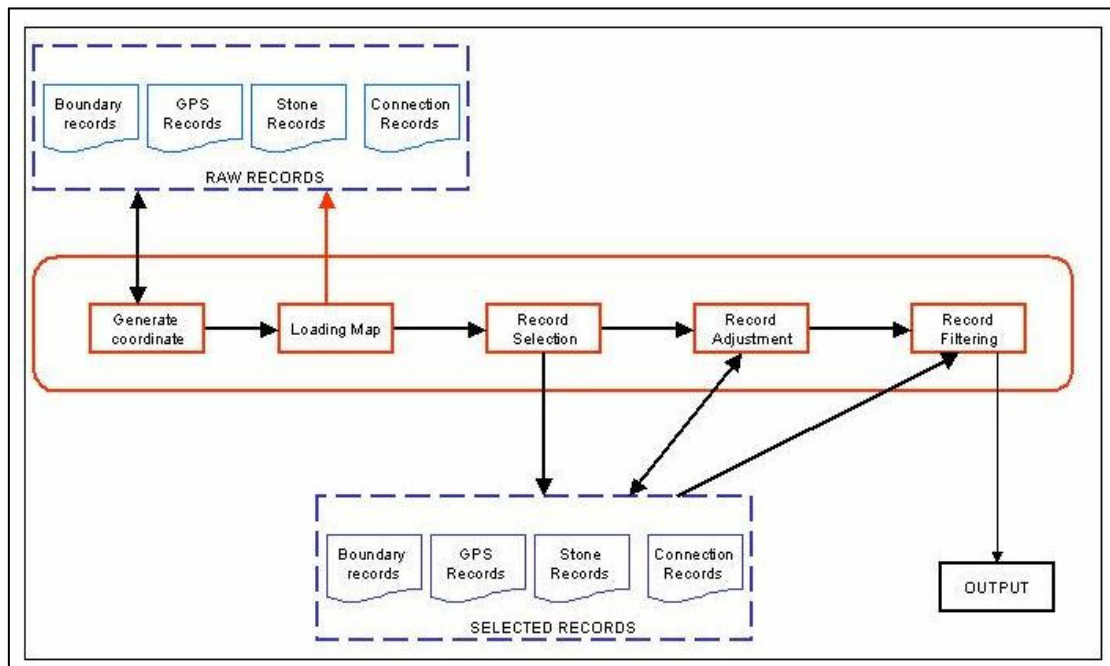
Figure 5. Overall Data Selection Unified Modelling Language for Prototype

3.3.3 Development of Artificial Intelligence Based Automated Cadastral Database Selection System

3.3.3.1 Identifying problem/errors and the solutions

In previous study (1996-2000), there are some main issues that related to DCDB and CCS requirement have been identified. All this issues will be stored as the knowledge base with its solutions method. With an automation mechanism, the system will expert enough to identify the problem and searching the possible solutions to solve the problem

3.3.3.2 Closed Polygon Recognition

In this system features, it includes the capability to identify a closed polygon in order to have a quality data for further adjustment. Closed polygon is important element in adjustment stage. This can be defined as one of intelligence mechanism because to identify one closed polygon requires a human knowledge.

3.3.3.3 Data Error Correction

Identifying data error and how to solve it is also one of the characteristics of intelligence. At initial stage, there some error occurs at existing database due to human error and technical problem. By using intelligence mechanism, the system could used to identify the data error, the type of data error and can recognize the method to correct the data error without affecting the quality of data.

3.3.3.4 Manipulation of information from different tables.

Inconsistency of data in different tables also can be solved by using an intelligence implementation mechanism. Sometimes the mark identities in each table are different even each point is having a same location. By using intelligence, the system systematically and automatically identifies the problems by searching the best method of corrections in various possibilities

3.3.3.5 Spatial and attribute computation based on survey concepts and Mathematical model

Outliers and errors in the database must be corrected in order to provide a good data input for adjustment process. All computations will be done automatically and the use of survey mathematical model is to ensure the reliability of the cadastral records in the database.

3.3.3.6 Automatic Weighting Strategy

In adjustment process, each observation required its own weight age. This information is needed for least squares technique to execute. This application will assign an individual weight to each observation that based on its class of survey. For bearing component, 10" and 30 "is assigned for First Class Survey and Second Class Survey, respectively. While 0.001 m and 0.010 m for distance component.
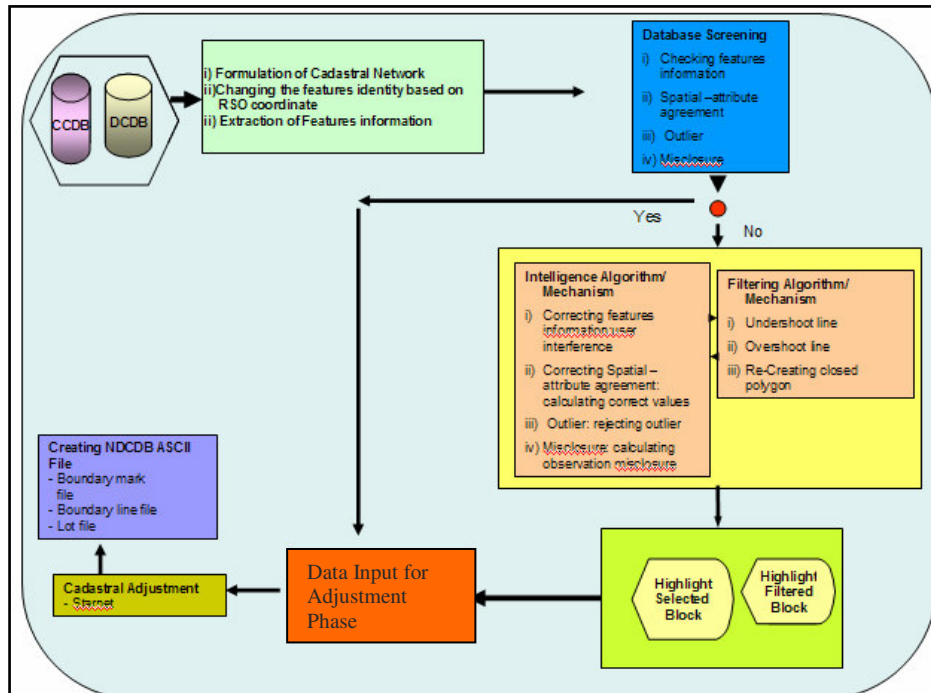
Figure 6. Conceptual of Intelligence System

4.0 CADASTRAL SELECTION AND VISUALIZATION PROTOTYPE

The prototype system has been developed in Windows environment by integrating the Visual Basic (VB) and application developer tools (ESRI Map Object and MapInfo Map X). An intelligent prototype was followed the conceptual model that has been outlined to handle different characteristics of DCDB. Figure 7 (a), (b), and (c) shows the physical model of the application.
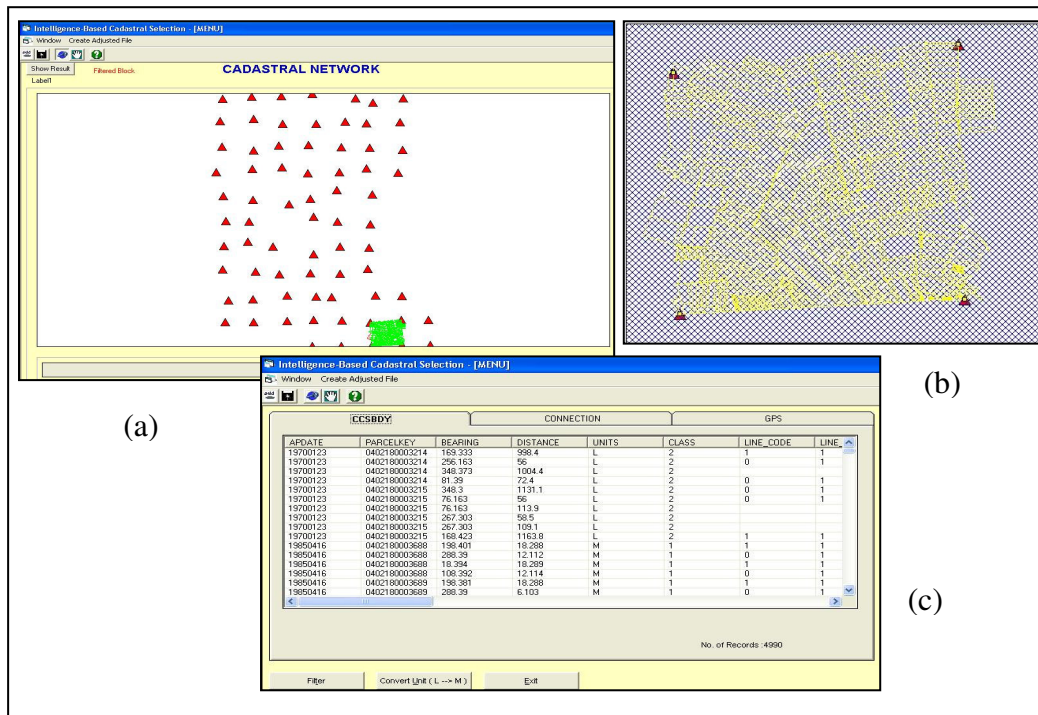

Figure 7. (a) Main Window (b) Cadastral Selection Process, and (c) Cadastral Database

Figure 7 (a), shows the main page of the application, which are consist of two menus; i) Map Visualization and ii) Database for particular spatial features. The process of cadastral block selection involves identifying control station and related DCDB as shown in Figure 7 (b). Boundary line, connection line, and cadastral control identities (From node and To node) will be converted to Rectified Skew Orthomorphic (RSO) coordinate system. RSO coordinate was adopted as an identity system because RSO provides a homogenous system for identification process. Information for the selected block is listed in database format. The application will conducts a database screening process in order to check the integrity of the data input. Based on the data screening result, the application is able to provide two options for a user; i) proceed with a Starnet data format conversion process, and ii) executes intelligence mechanisms for database self-correction steps (Figure 7 (c). Intelligence mechanisms were explained in Section 3.3.3.

One of the CCS implementation goals is to create a survey accurate cadastral database (NDCDB). In order to realize the goal, the application has provides "tools" for NDCDB generation as shown in Figure. The NDCDB process starts when cadastral network adjustments have been performed, where output form adjustment software will communicates with the application. The intelligence application provides two menus for NDCDB creation purposes as follows; i) creating adjusted boundary mark file (Figure 8(a)) and ii) creating boundary line file (Figure 8 (b)).

(a)   (b)
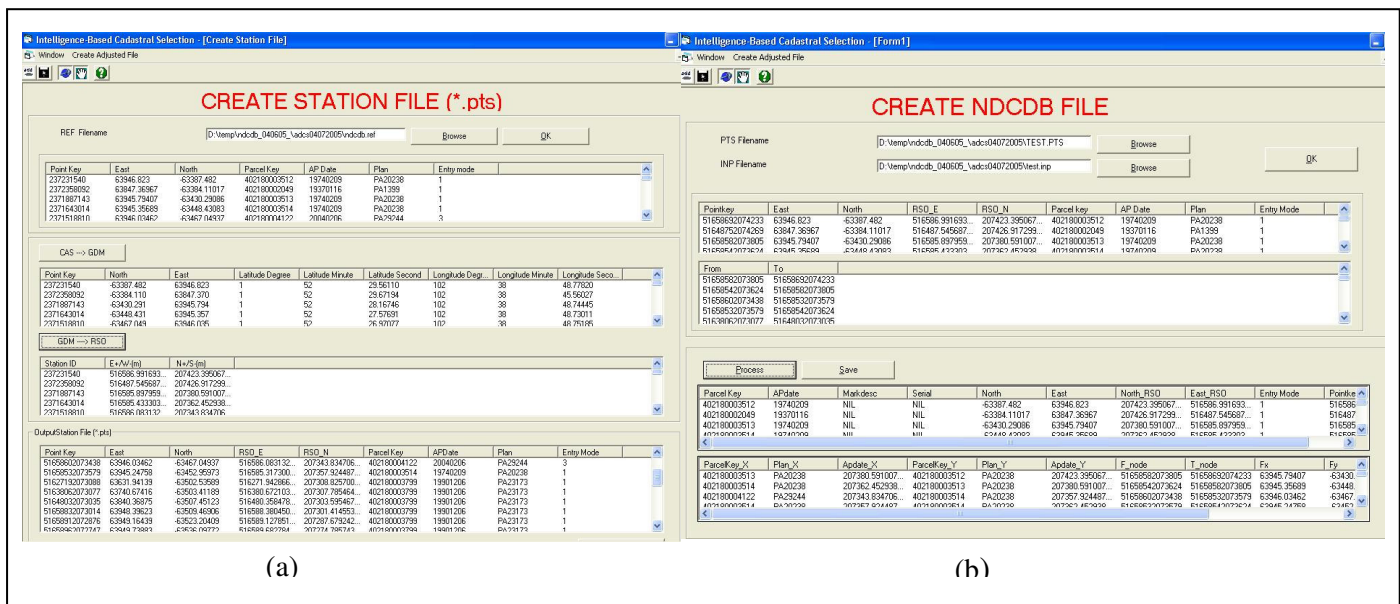
Figure 8. (a) Adjusted Boundary Mark Menu, and (b) Boundary Mark Menu

Current cadastral database (DCDB) was designed for many applications such as land title production process, engineering, GIS, planning, and decision-making purposes. Most of the applications need graphic and attribute information provided by DCDB. As a result, a proposed survey accurate DCDB (NDCDB) will also consists of graphic and attribute information as a current DCDB. Figure 9 visualizes samples of created survey accurate DCDB (NDCDB).
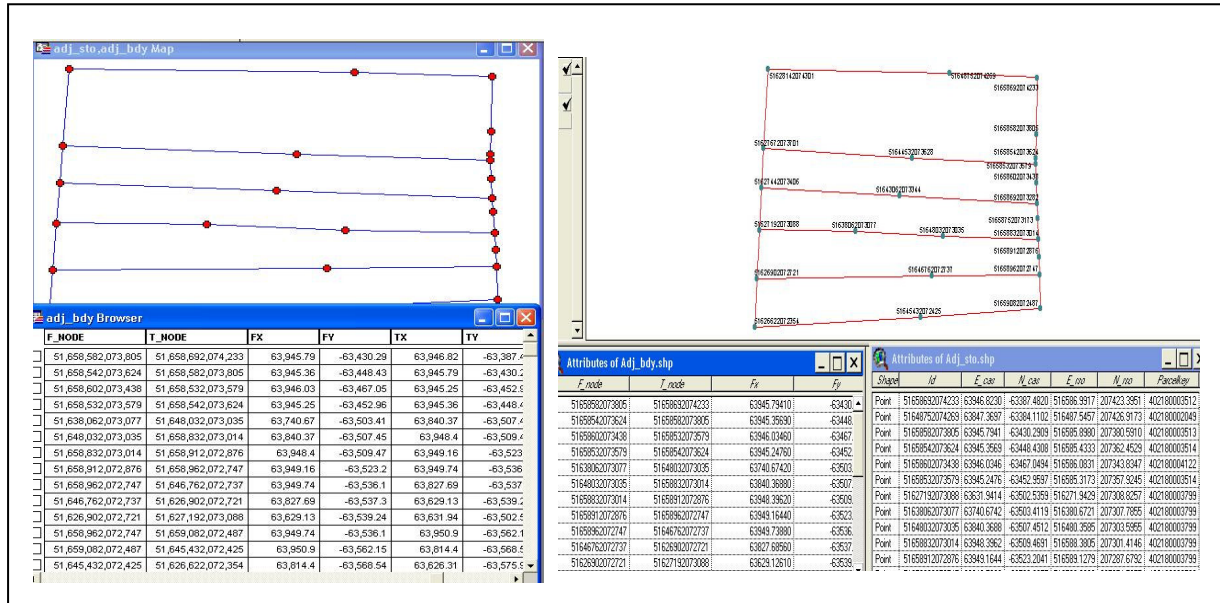
Figure 9. (a) Example of NDCDB (Boundary Line and Boundary Mark Files)

5.0   CONCLUSION

The development of the Intelligence Based Cadastral Database Selection and Visualization System will ease the NDCDB development process. Data screening and cleaning is essential since outliers exist in the data input. Intelligence and automated self-correction mechanisms will execute during the data input screening and filtering process and it will minimize manual editing process. The application will handle the incomplete spatial features and solve the outliers in the DCDB. Execution time for data selection, screening, and filtering depend on several factors, as follows:

        i)   Outliers
        ii)   Algorithm
        ii)   Block size
        iii) Number of boundary mark
        vi) Density of the cadastral lot

The result form the adjustment and new survey accurate NDCDB show that the Intelligence Based Cadastral Database Selection and Visualization prototype was functioning well and having a good reliability and integrity.

6.0   REFERENCES AND BIBLIOGRAPHIES

1.   Ahmad Fauzi Nordin (2002). *Institutional Issues On The Implementation of The Coordinated cadastral System For Peninsular Malaysia: A Study on the legal and Organisational Aspects*. Masters Degree Thesis.Universiti Teknologi Malaysia, Skudai , Malaysia

2.   Abdul Majid Mohamed, Chia Wee Tong and Chan Hun Seok (1998). *Cadastral Reforms in Malaysia*. FIG XXI Congress Proceeding, Commission 7. Brighton, United Kingdom.

3.   Abd Majid A. Kadir and Shahrum Ses (1999). *Report on Contract Research for Feasibility Study on a Coordinated Cadastral System for Malaysia: The Adjustment of Large Cadastral Network with Reference to RSO Coordinate System*. Department of Survey and Mapping, Malaysia. Kuala Lumpur

4.   Abd Majid Kadir and Ghazali Desa (2002). *Report On The Institutional Issues On The Implementation of Coordinated Cadastral System.* Department of Survey and Mapping, Malaysia. Kuala Lumpur

5.   Abdullah Hisam Omar (2004).*Development of A Coordinated Cadastral System For Peninsular   Malaysia*. Doctor of Philosophy Thesis. Universiti Teknologi Malaysia, Skudai, Malaysia.

6.   Russell,S and Norvig, P (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall: Berkeley.

7.   Winston,P.H (1992). *Artificial Intelligence.* Addison-Wesley; New Jersey.