The Journal of GIS Association of Korea, Vol. 14, No. 4, pp. 403-419, December 2006

A Geocoding Method Implemented for Hierarchical Areal Addressing System in Korea

Jiyeong Lee* · Hag-Yeol Kim**

ABSTRACT

The well-known address matching technology developed by the U.S. Census Bureau was applicable only for street addresses. However, many other addressing systems on the basis of a hierarchy of areas (hierarchical areal addressing system), such as Korean or Japanese addressing system, cannot be suitable for the existing address matching method. This paper, therefore, develops an areal address matching method, especially for Korean addressing system, in order to geocode 2D and 3D locational data of human activities. Thus, this study explains a new approach to dealing with 3D positioning method composed of two geocoding methods, which are a '2D Korean Address Matching' technique and a '3D Address Matching' technique.

Keywords : geocoding, hierarchical address system, address matching

요 약

현재 주소정합(address matching) 기술 중 가장 잘 알려진 방법은 미국 센서스국에서 개발된 것으로서 거리이름을 기반으로 한 주소체계에 그 적용이 가능하다. 그러나 한국, 일본과 같은 많은 국가들의 주소체계는 장소의 계층적 구조에 기초한 계층적 주소체계로 인하여 현재의 주소정합 기술을 적용하기가 불가능한 상황이다. 따라서 본 연구는 인간활동에서 생성되는 2차원적 · 3차원적 주소정보를 주소좌표변환(geocoding) 하기위한 지역적 주소정합방법, 특히 한국의 주소체계에 적합한 주소정합방법을 이론적으로 개발하는데 그 목적을 둔다. 이에 이 논문은 2차원 및 3차원 주소정합기술로 구성된 3차원 주소위치확인에 적용할 수 있는 새로운 접근방법을 설명한다.

주요어 : 주소좌표변환, 계층적 주소체계, 주소정합

* Assistant Professor, Department of Geography and Earth Sciences, The University of North Carolina at Charlotte 9201 University City Blvd., Charlotte, NC 28223, USA (jlee68@uncc.edu)

** Assistant Professor, Department of Urban Engineering, Seokyeong University 16-1 Jungneung-dong Sungbuk-ku, Seoul, Korea (hagkim@skuniv.ac.kr)

1. Introduction

Many studies about individual human activity and movement in the social sciences have required the detailed individual-level activity data. They cover a wide range of topics, including travel behavior, migration and residential mobility, pedestrian choice behavior, shopping and commuting behavior as well as Location-Based Services (LBS). Individual-level activity data also allow researchers to examine the complex interaction between space and time and its effects on the structure of human activity patterns in particular areas (Cullen, Godson and Major 1972). Recently, many studies have been conducted using individual activity data in geographic information systems (GIS) (Kwan 1998; Kwan 2000; Kwan and Lee 2004; Miller 1999; Weber and Kwan 2002). These studies are accompanied with generating and managing detailed individual-level activity data as well as implementing analytical tools in GIS (Kwan and Lee 2004; Miller 2004).

One of the most popular methods to generate accurate individual activity data in GIS is a geo-coding method, called an address matching method (Drummond 1995). It converts an street-address database collected by public or private sectors into a GIS database containing locational information. Since address data provide the linkage to mapping any geo-database, various GIS applications require the function to match an address to a geographic location. Using the address matching technique in GIS, potential spatial data analyses will be implemented in many public services such as E-911 services, public safety analyses, building permits maintaining and tracking, urban facility management, economic analyses, hazardous material inventories, real estate transactions, property ownership inventories, and not limited.

The well-known address matching technology, initially developed by the U.S. Census Bureau, has been available for more than forty years in order to allocate population accurately within blocks, census tracts, and other geographic areas, without expense of sending enumerators to every dwelling unit in the country (Drummond 1995). The address matching method, however, was developed for a network-based addressing system, called street addresses, such as the U.S. and European addressing systems. There are other addressing systems in the world, such as Korean and Japanese addressing systems, for instance. The Korean Addressing System was developed on the basis of a hierarchy of areas (Davis, Fonseca and Borges 2003), called 'hierarchical areal addressing system'. In order to geocode human activity data based on different addressing systems, a new address matching method suitable for the hierarchical areal addressing system should be developed and implemented.

Therefore, the purpose of this paper is to develop an 'hierarchical areal address matching' method, especially for Korean addressing system, in order to identify 2D and 3D locational data for analyzing human activities. Thus, a new approach to dealing with 2D and 3D positioning methods will be explained; a '2D Korean Address Matching' technique and a '3D indoor geo-coding' technique (Lee, 2004b).

Section 2 of this paper introduces current addressing systems and analyzes possibilities and limitations of geo-coding methods. Section 3 proposes a new 2D areal address matching method and a 3D indoor positioning method utilizing a conventional address matching technology. The following section describes the output from the experimental implementation of the 3D address matching technique and demonstrates the potential benefits of the address matching technique for improving the speed of emergency response. The final section discusses several substantive insights derived from this study.

2. Addressing Systems and Geocoding Methods

2.1 Addressing Systems

One source of geospatial information on administrative or business location is from postal addresses, which have been used to define a specific location around the world. However, standard or format of address data varies considerably from one country to another. The address formats can be classified into two systems. The first is a street network-based address system. It usually is called a street addressing system, used in the US and Western Europe. The other is a hierarchical areal address system, and used in Asia, especially Korea and Japan, employing a more cadastral characteristic, such as parcel or block. First, look over the former addressing system in detail. The general formats of the street addressing system are as follows:

Street Address (Single Housing Unit: 2D)
= House # | Street | City | State | Zip Code
Street Address (Apt. Complex: 3D)
= Room # | Building Name | House # | Street | City | State | Zip Code

The first addressing initiatives took place in Western Europe and China in the 18th Century (Davis, Fonseca and Borges 2003). While there are various addressing systems throughout the world, few standardized addressing systems are developed and implemented in the developed countries. One of the most important street addressing systems is the metric numbering system, combined with odd-even rules. Any building or house is assigned a number based on its metric distance from the beginning of the street segment. The number is rounded up to the nearest odd or even number, or approximated in a way that every building gets a unique number. There are many ways in which the numbering is sequential and block-oriented, but the numbering is not distance-oriented. The both numbering systems have the advantage of allowing an easy assessment of the distance between two addresses in the same street, but the sequential addressing, sequential odd and even numbers at each side of the streets, fails to assign new intermediate numbers for new buildings or re-partitioned parcels. In the United States, the block-oriented numbering system is widely used to assign the building or parcel numbers based on a proportion of addresses arranged to the block segment distance (Kim 2001). The block addressing information is coded as a set of attributes for segments of street centerlines in TIGER (Topologically Integrated Geographic Encoding and Referencing) files (Drummond 1995).

In order to assign addresses to apartment complexes and commercial buildings in urban areas, a building with multiple apartment or suite addresses typically uses the extended street addresses, called '3D Street Addresses'. As seen above, the 3D address have a little different address format, called "a composite address", consisting of two components, an indoor address and a street address. The indoor addresses such as the suite, room and apartment numbers represents the locations of each of the building's interior floors or rooms on a floor. The indoor addresses can support local government services such as emergency services, code enforcement, taxation, building ownerships, crime reports, delivery services, etc. that refer to the 3D spaces within a building. For example, the address of a lawyer's office located in Minneapolis, MN is "Suite 300, 701 Fourth Avenue South, Minneapolis, Minnesota 55415-1810." "Suite 300" represents the indoor address of the office, while the rest, "701 Fourth Avenue South, Minneapolis, Minnesota 55415-1810," represents the street address.

The other addressing system is a 'hierarchical areal address', a hierarchical area-based address system. Korean addressing system is a kind of the area-based address systems, which use the concepts of a more cadastral nature, such as parcel or block as address references. The general formats of the Korean addressing system in urban areas are as follows:

Area-based Address = City | Ward | Town | District | Neighborhood | Block | House # | Postal Code

Area-based Address (3D) = City | Ward | Town | Parcel # | Building or Apartment Name | Building # | Apart or Suite # | Postal Code

The Korean addressing system is based on hierarchical areas, subdividing an area into small places. There are different types of subdivisions in urban and rural areas. As seen in Figure 1, the country is divided into metropolises ("Shi" in Korean) and Provinces ("Do"). The provinces are divided into counties ("Kun") and cities ("Shi"). Small cities are divided into districts or neighborhoods ("Dong"). Geographic hierarchy in metropolises is wards ("Ku"), districts ("Dong"), neighborhoods ("Tong") and blocks ("Ban"). There are, in general, between fifteen and twenty house units within a block. The buildings within a block were

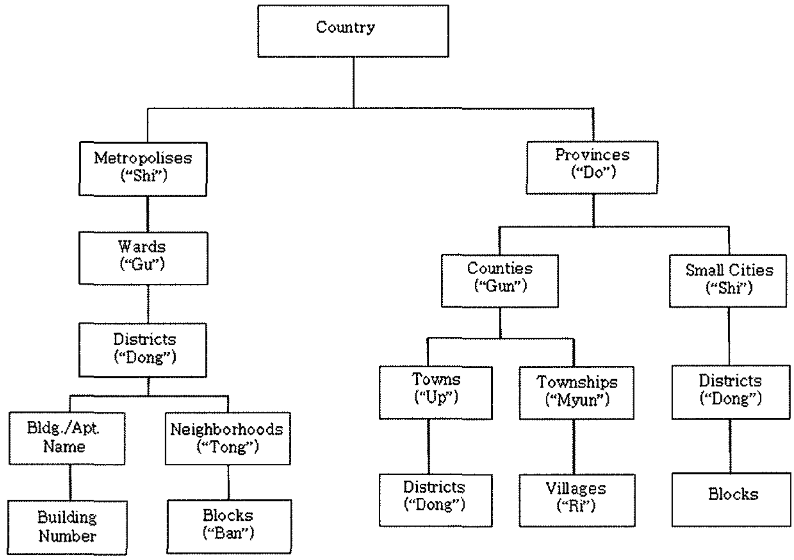


Figure 1. Geographic Hierarchy for Korean Addressing System

initially numbered in the order of a northwest numbering rule, starting from number 1 by the direction of northwest to southwest based on the order that the buildings were built (Kim 2001). In Korea, the sequential numbering scheme fails to assign new intermediate numbers for new buildings or re-partitioned parcels, and the divided parcels are numbered by attaching the sub number to an original number. This is the parcel or building number before partition, such as 101-2 (Kim 2001). Multiple apartment or suite addresses for a building typically use 3D indoor addressing systems (Lee 2004b).

2.2 Geocoding Methods

Geocoding is a process of placing geographic locations (identifiers) for tabular data. One of the most popular geocoding methods is an address matching in GIS. Geocoding is performed as a preliminary step to spatial analysis of point data.

The address matching process includes three steps: a parsing step is to deal with semi-structured input addresses of events, a matching step is to make a correspondence between the input address and the addressing reference database, and a locating step is to assign coordinates of the event, as seen in Figure 2. In order to perform the three steps, the geocoding process requires two components: a reference database in which information about the addressing system are stored, and geocoding methods which will be performed in several ways depending on the available reference address databases.

The parsing step converts the addresses of

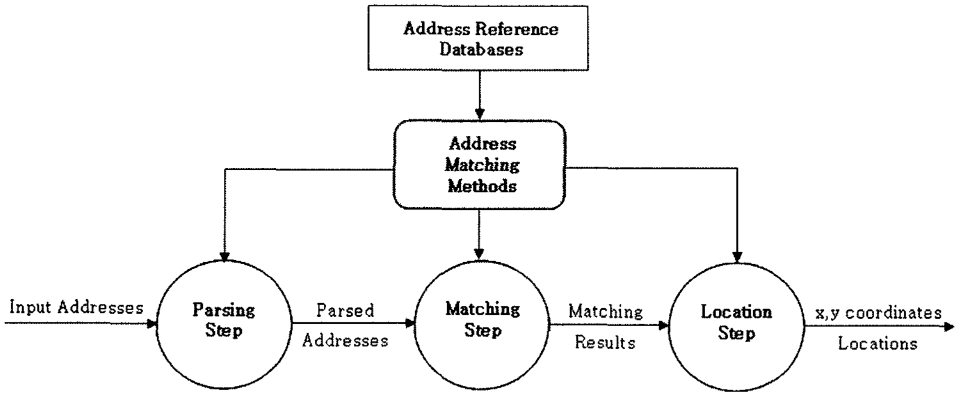


Figure 2. Procedure of Address Matching

an input database into a structured and standardized format for a matching step. In the address field of the input database, major problems include abbreviated street names, misspelled street names, complex directional combinations, and ambiguous information. The parsing step can be done through various text- and string-oriented algorithms (Davis, Fonseca and Borges 2003) using translation tables and the Soundex function (Knuth 1973). The matching step tries to find the most precise reference information from the reference address database, based on the given input address. Depending upon the reference address databases, the matching techniques are implemented in several ways.

Two kinds of reference address database have been used in current address matching methods: street centerlines with address ranges (US Census 1993) and point-georeferenced individual addresses (Ordnance Survey 2004). With using the street centerlines with address ranges as the reference database, for instance, the TIGER/Line files developed by the U.S. Census Bureau, the matching step identifies a street-segment record in the reference database that has the same street name, street type, and other identifiers as the record in the input addresses. After the two records have been matched, interpolation is used to assign geographic coordinates to the input addresses in the locating step (Walls 2003).

In the United Kingdom, the Ordnance Survey produces an Address Point database, called Address Layer, that provides precise coordinates for more than 26 million residential and commercial properties in Great Britain. This database provides for the most accurate and up-to-date link between any property address and its location on the map (Ordnance Survey 2004). Through the Address Layer, the matching step establishes precise links between input addressable properties and all geographic information in the reference database.

Finally, the locating step determines the actual geographic coordinates (x and y coordinates), assigned to the input addresses. Depending upon the reference address database, the locating step either copies the coordinates of a point object in the Individual point Address database, or interpolates to determine the address's approximate distance along the street segment by using the low and high addresses of the block face as the well-known street address matching techniques (US Census 1993).

However, since most current geocoding methods have been developed to determine the locations of street addresses, they cannot be used for the area-based address system. Therefore, this paper proposes to develop a new '2D and 3D Korean address matching' method.

3. Address Matching Method for Korean Addressing System

To implement a Korean address matching function requires two procedures. One is to model korean addressing mechanism to make georeference files. The other is to implement an "Korean Address Matching" function. A database schema for the Korean addressing system database is proposed in Figure 3. Although the Korean addressing system has different types of sub-divisions depending on urban or rural areas, the database schema has been designed only for the urban area addressing system in this paper. The schema considers the full range of addressing components, including municipality boundaries,

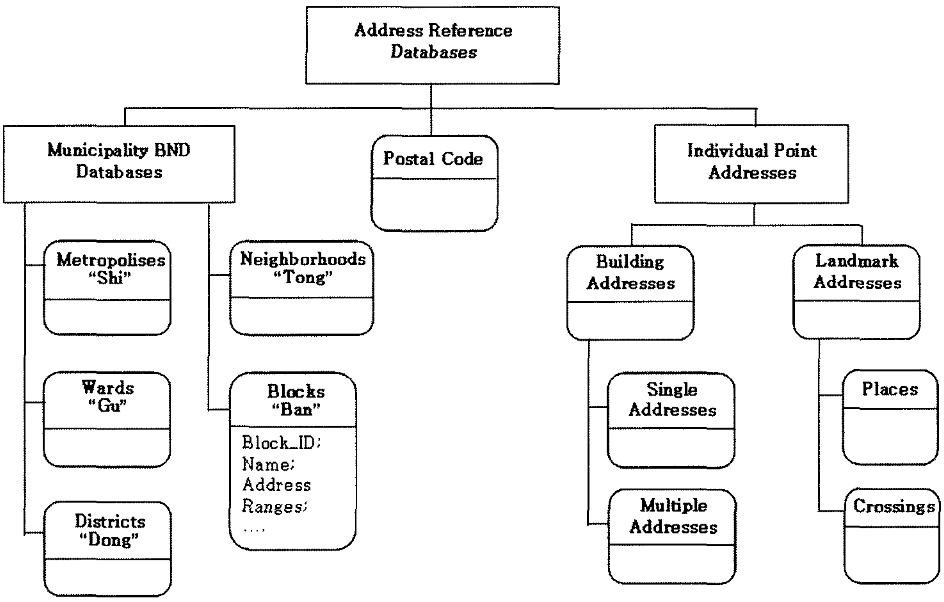


Figure 3. Conceptual Reference Database Schema

Postal-code boundaries, and reference point databases. Because the Korean addressing system is based on hierarchical areas, the municipalities are subdivided from big to small such as Metropolises ("Shi"), Wards ("Gu"), Districts ("Dong"), Neighborhoods ("Tong") and Blocks ("Ban"). The Block object has attributes including 'Block_Code', 'Name' and 'House Number Ranges'. The reference point database contains the Building Address class and the Landmark Point class.

Although several strategies can be implemented from such designed address databases, this paper focuses on developing a 2D Korean address matching method using the Municipality Boundaries database as a reference database, as well as a 3D Korean address matching method using the Building Address database.

3.1 2D Address Matching for Korean Addressing System

Let us suppose we have a following single house address in Korea. : "Seoul, Nowon-gu, Chang-dong, 225 10-tong 2-ban, 139-011". The last number is the postal code, and "225" is the building (house) number. "10" is the neighborhood number and "2" is the block number. In order to geo-code this Korean address, two steps are proposed: one is to geocode the address based on the areal address (e.g. "Seoul, Nowon-gu, Chang-dong, 10-tong 2-ban") to define the location information of the block, and the other is to define the positioning of the building within the block based on the building number (e.g. "225").

To define the house location within the block in the above example, the first step is to select the 'Nowon-gu' polygon from a 'Ward' layer by querying based on the attribute data (Ward_Code). Based on the selected polygon and a District layer, the districts within the 'Nowon-gu' polygon are selected by a spatial query ('Selection by Location' function in ArcGIS). Likewise, the 'Chang-dong' polygon is selected, and then the '10-tong' and '2-ban' polygon are selected. After querying the '2-ban' polygon by an attribute selection based on a Block_Code, a building needs to be geocoded within the block ('2-ban') based on the building number ('225') and the "2D Geocoding" method.

The current area-based geocoding methods to identify the location information of individual events within the polygon are 1) the generation of a random point inside the area (Block) boundaries (Davis, Fonseca and Borges 2003) or 2) an area centroid geo-coding method (ESRI 2001). However, this paper proposes a new way to geocode the building addresses within the block boundaries, called '2D Korean Address Matching' method.

Since a polygon is described with connected lines in the vector data model for geometric representations, the block polygon structure is identical to the polyline structure (ESRI 1998). A polygon is a connected sequence of three or more points that form a closed, non-self-intersecting loop. The primary classes of the Block Data Model (BDM) are Point and

```
class Point {                     class Block {
        Int       Point_ID;       Int       Block_ID;
        Double    x, y;           Int       NumPoints;
        };                        Int       FHouseNum;
                                  Int       THouseNum;
                                  Point[NumPoints]     Points;
                                  };
```

Figure 4. A Block Data Model

Block, whose schema is shown in Figure 4. The block structure consists of a set of Points *pt* and other attributes including an identifier, and total number of points. In addition, two attributes store a starting house (or building) number associated with the first point of the set of Points (*Point[0]*) and an ending house number within the block, which information will be used for 2D indoor geocoding explained in the following section. Points for the block polygon are always in clockwise order and the first and last point of the block must be the same.

The fundamental principle of the 2D address matching method for areal addresses employs a traditional street address matching technology developed for location positioning of outdoor environments. A block in the model contains building numbers as a range attribute, which refers to a starting house number (FHouseNum) on the first point of the block and an ending house number (THouseNum) on the end point of the block. In general, each house within the block ('2-ban') is uniquely labeled. The house label provides a useful value in a database table because the label is associated with parcel numbers in a Cadastre System (Lee 2004b). Unlike a street address, the house number (or parcel number) is not always assigned as a number in sequence, or as the "double parcel number" including a sub number such as '221-2'. In order to standardize the house numbers (using sequential numbering in clockwise order), translation tables can be used. For example, a translation table would change "212", "212-1" and "211" as the standard form into "211," "212," "213," respectively, before geocoding the house address within the block. After defining the house location based on the standardized house number, the georeferenced point will contain the original house address (input address) by simply copying the un-translated housing address.

In order to identify the location of House *k* in the block model (BDM) for a Block *i*, the 2D address matching method will be executed in four steps. The entire procedure for geocoding the parsed house address within the block is summarized as follows: step one is to calculate a network distance (d_{0k}) from a starting point (pt_0 = *Point[0]*) along the boundary of Block *i* (B_i) based on the target house number (k) and the housenumber ranges of blocks in BDM. Next Step is to identify a segment of the Block i (B_i) according to $d_{0j} < d_{0k} < d_{0j+1}$, where d_{0j} is a network distance from pt_0 to pt_j in the Block i (B_i). The third step is to determine the locational positioning (n_k) of *HouseNum_k* (with x, y coordinates). The final step is to apply a user-specified offset to move the point location (representing a house) a certain distance (inside) away from the boundary of Block i (B_i).

As seen in Figure 5, the distance d_{0k} is equal to the distance (d_{0t}) multiplied by the proportion of house number ranges ((*HouseNum_k* - *FHouse_Num*) / (*THouseNum* - *FHouseNum*)), when t = the total number of points (*NumPoints*) in B_i. The distance d_{0t} is the sum of all segments' lengths in the block polygon B_i. The next step identifies a segment of the Block i (B_i) according to $d_{0j} <= d_{0k} <= d_{0j+1}$, where d_{0j} is a network distance from pt_0 to pt_j in the Block i (B_i). Suppose $d_{02} < d_{0k} < d_{03}$, the segment s_{23} connecting pt_2 and pt_3 is identified, as well as the distance d_{2k} is calculated by d_{0k} - d_{02} in Figure 5. The segment s_{23} is characterized by a linear function $Ax + By + C = 0$. The x,y coordinates of a point pt_k (representing a House k) are calculated by the mathematical functions described in Figure 5 (Lee 2004a). V is a unit vector of the segment s23.

The final step is to apply a user-specified offset to move the point location (representing a house) a certain distance (inside) away from

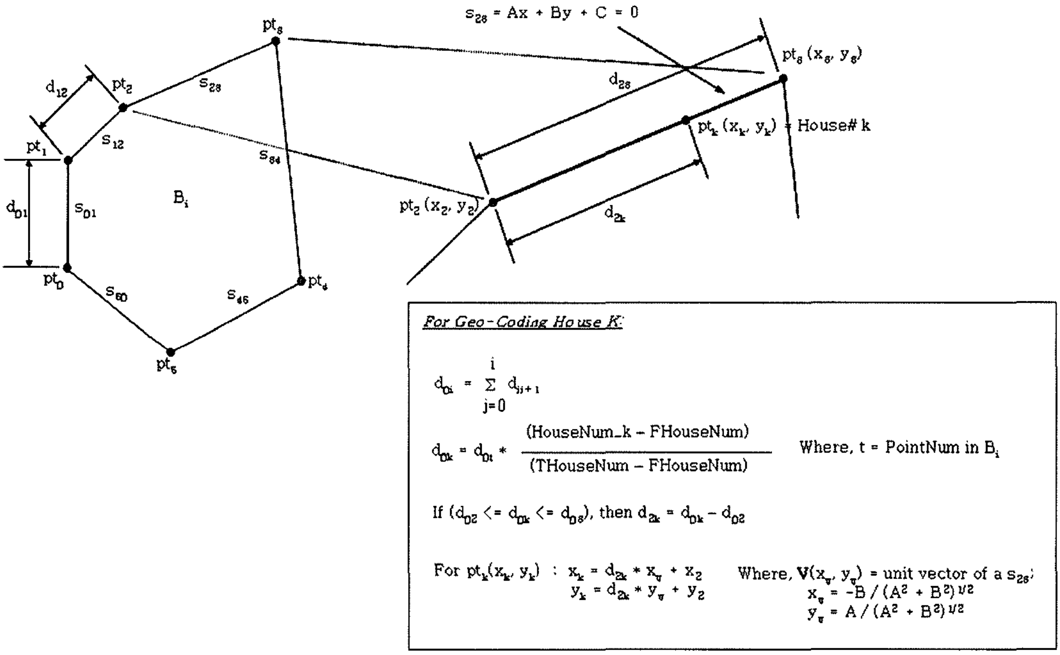


Figure 5. 2D Address Matching for Areal Addresses

the segment s_{23}. As seen in Figure 6, a vector method for identifying a right offset of the point pt_k is to calculate the edge-vector cross product (Lee 2004a). V_1 is a unit vector of a segment s_{23}, and V_2 is a unit vector being perpendicular to a segment s_{23} with a right direction. The cross product $V_1 \times V_2$ for two successive segment vectors is a vector perpendicular to the xy-plane, with a z-component equal to $-V_{1x}V_{2y} + V_{1y}V_{2x}$, which should be negative. In order to avoid working in 3D, the "dot-perpendicular" product is used, which is a signed measure of the area of the parallelogram determined by two vectors: $(V_{1x}, V_{1y}) \cdot (V_{2x}, V_{2y}) = (V_{1x}, V_{1y})$ $(-V_{2y}, V_{2x}) = -V_{1x}V_{2y} + V_{1y}V_{2x} = z$ coordinate of $(V_{1x}, V_{1y}, 0)$ $(V_{2x}, V_{2y}, 0)$ (Lee 2004a). The x,y-coordinates of the new point pt_k (representing a House k) with a right offset are calculated by the functions described in the previous step (Figure 5).

3.2 3D Address Matching for a Korean Addressing System

The 2D address matching method proposed in the previous section is to determine approximate locations of the given database, as same as the street address matching method does. One of the most precise methods is the matching of the input address to an object from the individual point address reference database as implemented in ADDRESS-POINT. Ordnance Survey, Inc. in the UK developed the ADDRESS-POINT application to define and locate residential, business, and public postal addresses in Great Britain, based on more than 25 million addresses recorded in Royal Mail's Postcode Address File (PAF®).

Each point in the Building Address database may represent a building, a single family house or an apartment complex / commercial building. Thus points should be classified into two different classes, depending on that

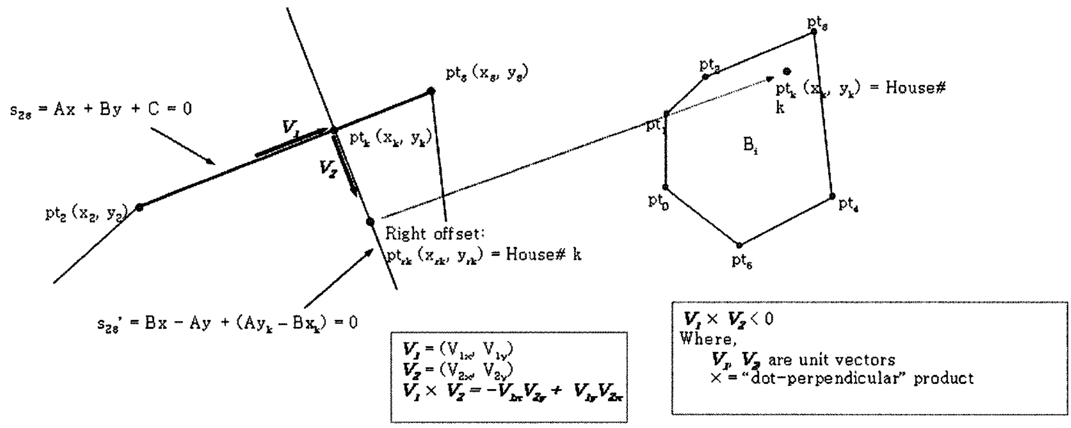


Figure 6. Right Offset for 2D Address Matching

a point represents a single family house or an apartment complex. The Building Address database contains both the Single Address class, called a Point, and the Multi-Address class, called a Master_Point in a hierarchical structure (Mainguenaud 1995). The Master_ Point is an abstract representation of the internal structure of a building, which includes attribute data on multiple apartment units or suite addresses of the building. For example, if there is an apartment unit address in Seoul, "Seoul, Kangnam-gu, Banpo-dong, 1319 Jukong Apt. 209-dong 804, 129-011", the Master_Point means "Seoul, Kangnam-gu, Banpo-dong, 1319 Jukong Apt. 209-dong".

In order to model the internal structure of a building to serve as reference data for geocoding individual apartment or suite address in 3D, Lee (2001, 2004a) has proposed a 3D Geometric Network Model (GNM), which describes a topological model for representing connectivity relationships between 3D objects. The model is based on a Node-Relation Structure (NRS) (Lee 2001). Such a 3D network data model is described based upon the work of Lee (2001). Instead of describing all internal structures of a building (including all rooms), the model represents the transport routes using 3D Geometric Network Model (GNM) as a reference data for 3D geocoding. The connectivity relationships among 3D spatial units (transport routes within a building) are the combinations of the connectivity relations both in the horizontal directions on a floor and in the vertical direction among floors. The horizontal connectivity relationships on a floor are derived from the 2-dimensional connection among polygons. The vertical connectivity relationships are defined by the locations of stairways or elevators.

In order to model the connectivity relationships among the transport routes on floor j, 3D Poincaré Duality Transformation (Lee 2001) is utilized to define the dual graph of 3D objects. The connectivity relationships are modeled as a 3D network structure, the dual graph of 3D spatial units. A graph $G = (V(G), E(G))$ consists of two sets: a finite set V of elements called vertices and a finite set E of elements called edges, $V(G) = \{v_1, \cdots\cdots, v_n\}$ and $E(G) = \{(v_i, v_j) \mid v_i, v_j \in V\}$. That is, each edge is identified with a pair of vertices (v_1, v_2), where $v_1, v_2 \in V$ (Mainguenaud 1995).

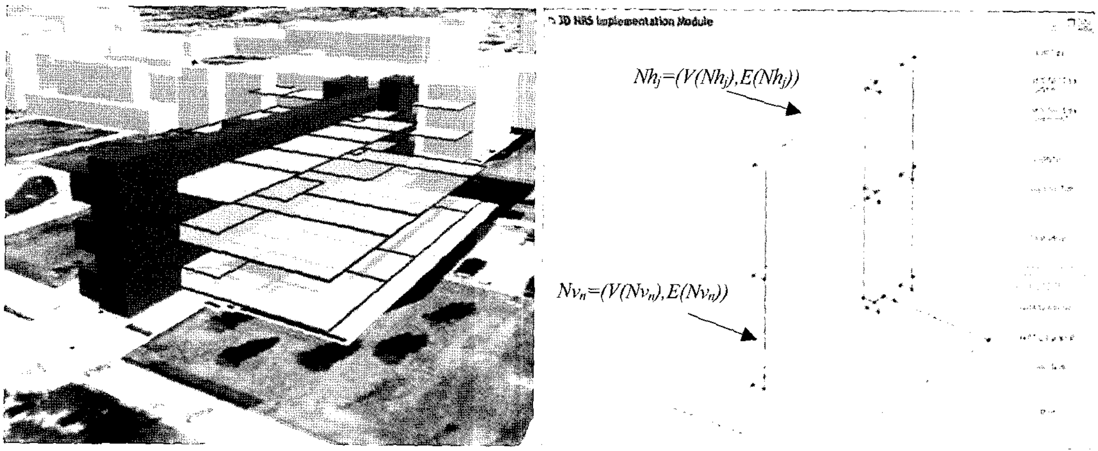
To represent the transport routes using the 3D GNM, a graph $Nh_j = (V(Nh_j), E(Nh_j))$, the first step is that a hallway is transformed into a linear feature (Blum 1967) based upon the Straight Medial Axis Transformation (MAT), a graph $MAT_j = (V(MAT_j), E(MAT_j))$ (Lee D.T. 1982; Aichholzerand Aurenhammer 1998; Lee 2004a). Next, each node representing 3D spatial units (stairways or elevators) is projected and connected into the medial axis. The projection $p(q, e)$ of a point q onto an edge e is determined by the intersection of the edge e and the edge perpendicular to the edge e through the point q (Lee 2004a).

The graph $Nh_j = (V(Nh_j), E(Nh_j))$, representing the geometric network within floor j (Figure

7b), is combined with graph $Nv_n = (V(Nv_n), E(Nv_n))$ using a union operation to produce the graph $N_i = (V(N_i), E(N_i))$ for a building i (Figure 7b), which is the geometric network model of the NRS. The graph $Nv_n = (V(Nv_n), E(Nv_n))$ is a subgraph representing the connectivity relations in vertical directions between floor j-1 and floor j. Figure 7a displays the transport routes within the study building in 3D View, and Figure 7b depicts the extracted 3D Geometric Network Data Model of the study area. The figure presents the connectivity relationships among transport routes (such as corridors, stairwaysand elevators) inside the building.

The network structure of a building, a graph $N_i = (V(N_i), E(N_i))$, consists of a set of Nodes V and a set of Edges E. The primary classes of the model are Node, Edge, Network and Master_Point, whose schema is shown in Figure 3. The class Node consists of an identifier, transport route types and position data in 3D (x,y,z-coordinates). The class Edge consists of an identifier, start node, end node, and four apartment or suite address ranges, which are the low and high apartment or suite numbers on the edge's left side and the low and high apartment number on its right side, which information will be used for 3D address matching. The class Network consists of an identifier and lists of all nodes and of all edges in a network. The Master_Point in the database has an identifier, a building address, a building name (or building number), and a 3D network representing an abstract of the internal structure of a building and including attribute data on multiple apartment or suite addresses of the building.

The 3D Address Matching for Korean Address System is based on Lee's 3D Indoor Geo-Coding method (2004b). A TIGER-type reference database model can be comprised of the 3D



(a) Transport Routes within a Building

(b) 3D Geometric Network Model

Figure 7. Modeling Internal Structure of a Building (Lee 2004b)

```
class Node {
        Int         Node_ID;
        Double      x, y, z;
        String      TransportType
        };

class Network {
        Int         Network_ID;
        Node        ArrayNode = new Node[];
        Edge        ArrayEdge = new Edge[];
        };

class Edge {
        Int         Edge_ID;
        Float       Length;
        Node        initial_node;
        Node        end_node;
        Int         F_UA_l
        Int         T_UA_l
        Int         F_UA_r
        Int         T_UA_r
        };

class Master_Point {
        Int         MPoint_ID;
        String      Areal_Address       // Address for building
        String      BuildingNum         // Building Name
        Network     Bldg_Network = new Network[];
        };
```

Figure 8. A 3D Network Data Model

GNM. An edge in the 3D GNM, a hallway line segment, contains an apartment or suite address range, which refers to the low (F_UA_l) and high (T_UA_l) unit numbers on the edge's left side and the low (F_UA_r) and high (T_UA_r) unit number on its right side.

In general, each unit in a building is uniquely named or labeled. The unit name or label provides a useful value in a database table as well as an intuitive indicator of the floor. For example, Suite 200 implies a specific place on the second floor of a building. In order to identify the location of the apartment unit k in the indoor network (3D GNM) for building i, $N_i = (V(N_i), E(N_i))$, the 3D Korean Address Matching method is executed in four steps. A short procedure is as follows: step one is to query a specific edge (e_l) defining the target unit number (*Unit_k*) by comparing it with the unit number ranges of edges in 3D GNM. The next step is to calculate a distance (d_{lk}) from From_Node (FN_l), and the following step is to determine the location positioning (n_k) of *Unit_k* (with x, y, z coordinates). The final step is to apply a user-specified offset to move the node location (representing a room) a certain distance away from the hallway line segment. The detail information on 3D Indoor Geo-Coding method is found in the Lee's paper (2004b).

4. Conclusions

Currently in Korea, there are different addressing structure among rural area address, urban area address, single family house address, and apartment or suite address as discussed above. In order to maintain the address data consistently in databases, the geocoding system requires the development of database schema for the address system. In order to resolve the problem, a government may wish to systemize its address scheme by reassigning numerical addresses and modifying street name, or by transforming to other address systems (Eichelberger 1993).

In this context, this paper developed an 'Korean Address Matching' method for Korean addressing system, in order to identify 2D and 3D locational data for analyzing human activities in urban environments. It presented a new approach to dealing with 2D and 3D positioning method composed of two address matching methods, which are a '2D Korean Address Matching' technique and a '3D Korean Address Matching' technique based on the 3D Indoor Geocoding technique (Lee 2004b). Since most current geocoding methods have been developed to determine the locations of street addresses, they cannot be applicable for the Korean Addressing System

While focusing on developing address matching methods for Korean address system, this paper ignored several issues related to the implementation and deployment of the address matching techniques. The first issue is related to the parsing step of address matching methods. This paper disregarded designing the translation tables to transform unstructured address data to structured and standardized database. The next one is to develop the actual addressing infrastructure with object classes: point-georeferenced individual building addresses; block data with house number ranges; and Master_Point database representing the internal structure of a building. In Korea when trying to find a place, it is a good idea to ask the following types of questions: which subway is closest? Is it near a prominent landmark? (e.g. public buildings or local famous restaurants) (World English Service 2004). In other words, a location within a city is specified as "at the intersection between two roads" or "nearby the landmarks of the city". This situation is included in the conceptual database schema presented in this paper. Further research is therefore required to develop a geocoding method based on the landmarks.

Third, in order to integrate with a 3D visualization system to allow the manipulation and exploration of georeferenced virtual environments, an additional data model is needed for geometric representation to visualize geographic entities in 3D viewers. This means that dual data models are required for implementing the 3D address matching technique in a 3D GIS. In order to maintain the data consistency between two data models, 3D GISs require the development of database models. Finally, the comprehensive GIS data to implement the 3D address matching technique raise serious concerns about issues of data security, as the data can be misused by unauthorized people. Means for preventing access to the data by such people should be considered before employing the system.

References

Aichholzer, O. and F. Aurenhammer, 1998. Straight Skeletons for General Polygonal Figures in the Plane, In Cai, J. Y. and Wong C. K. (eds.) *Proceedings of the Second Annual International Conference, COCOON '96, Hong Kong, Computing and Combinatorics, Lecture Notes in Computer Science 1090*, Berlin, Springer-Verlag: 117-126.

Blum, H., 1967. A Transformation for Extracting New Description of Shape, *Symposium of Models for Perception of Speech and Visual Forms*, MIT Press, Cambridge, MA, 362-380.

Cullen, I., V. Godson, and S. Major, 1972. The structure of activity patterns. In: Wilson, A.G., (Ed.), *Patterns and Processes in Urban and Regional Systems*. Pion, London, 281-296.

Davis, C., F. Fonseca, and K. Borges, 2003. A Flexible Addressing System for Approximate Geocoding. *GeoInfo 2003*. Available online at http://www.geoinfo.info/geoinfo2003/papers/geoinfo2003-25.pdf.

Dijkstra, E. W., 1959. A Note on Two Problems in Connection with Graphs, *Numerical Mathematics*, Vol. 1: 269-271

Drummond, W. J., 1995. Address Matching: GIS Technology for Mapping Human Activity Patterns. *APA Journal* (Spring), 240-251.

Eichelberger, P., 1993. The Importance of Address: The Locus of GIS. In *Proceedings of the 1993 Conference of the Urban and Regional Information Systems Association* 1: 212-217.

ESRI, 1998. *ESRI Shapefile Technical Description: ESRI White Paper*. ESRI, Redland, CA. http://ordnancesurvey.co.uk.

ESRI, 2001. *ArcGIS User Guide*. ESRI Press, Redlands, CA.

Kim, U. N., 2001. A Historical Study on the Parcel Number and Numbering System in Korea. *In Proceedings of the International Conference of the International Federation of Surveyors*. Available online at http://www.fig.net/figtree/pub/proceedings/koea/full-papers/pdf/session10/kim.pdf.

Knuth, D. E., 1973. *The Art of Computer Programming, Volume3: Sorting and Searching*. Reading, Addison-Wesley, MA.

Kwan, M. P. and J. Lee, 2004. Geovisualization of human activity patterns using 3D GIS: A time-geographic approach. In M. Goodchild and D. Janelle, (eds.): *Spatially Integrated Social Science: Examples in Best Practice*. Oxford University Press, Oxford 48-66.

Kwan, M. P., 1998. Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework. *Geographical Analysis* 30(3): 191-216.

Kwan, M. P., 2000. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set. *Transportation Research C* 8: 185-203.

Lee, D. T., 1982. Medial Axis Transformation of a Planar Shape, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4 (4): 363-369.

Lee, J., 2001. A 3D Data Model for Representing

Topological Relationships Between Spatial Entities in Built-Environments, Ph.D. Dissertation, Department of Geography, The Ohio State University.

Lee, J., 2004a. A Spatial Access Oriented Implementation of a Topological Data Model for 3D Urban Entities, *GeoInformatica* 8(3): 235-262.

Lee, J., 2004b. 3D Urban GIS for Geo-coding Human Activity Patterns, *Springer's Lecture Notes in Computer Science Computers.*

Mainguenaud, M., 1995. Modeling the network component of geographical information systems. *International Journal of Geographical Information Systems* 9: 575-593.

Miller, H. J., 1999. Measuring space-time accessibility benefits within transportation networks: basic theory and computational procedures. *Geographical Analysis* 31(2): 187-212.

Miller, H. J., 2004. What about people in geographic information science? In: P. Fisher and D. Unwin (eds.) *Re-Presenting Geographic Information Systems*, John Wiley.

Ordnance Survey, 2004. *OSMasterMap Address Layer User Guide*, Available online.

U.S. Bureau of the Census, 1993. *TIGER/Line Census Files, 1992: Technical Documentation.* Washington DC: U.S. Department of Commerce.

Weber, J., and M. P. Kwan, 2002. Bringing time back in: A study on the influence of travel time variations and facility opening hours on individual accessibility. *Professional Geographer*, 54(2): 226-240.

Wells, M. D., 2002. In Consistency in Address Assignment still Needed? In *Proceedings of URISA 5th Annual GIS in Addressing Conference: Street Smart and Address Savvy*, Province, RI.

World English Service, 2004. Postal Addresses in Korea. Available online http://teachkoreanz.com/living/address.htm. Last access Sept. 16 2004.