

An Addressing Model for Three-Dimensional City Properties in Hong Kong

Lilian S.C. Pun-Cheng and Y.C. Lee

Abstract: *A good address system has always been important for the systematic recording of personal or company information and for the effective delivery of products. In a society that demands rapid exchanges of information, an address serves also as a key to associate various governmental records, to link data of varying levels of summary, and to assist in strategic planning. In a city with complex building structures and street networks such as Hong Kong, a number of problems are encountered with addresses. These include the existing diverse address systems retained in a number of governmental units; the bilingual nature of the address system caused by a legacy of Hong Kong's past as a British colony populated largely by the Chinese; the three-dimensional variation of defining spatial units; and the lack of a clear postal code system all have led to difficulties in establishing a unified address system for the exchange of information. In this paper, the authors will try to examine all of these complexities in terms of the historical development of the address system and the physical setting of the city, followed by a proposal for a compartmentalized addressing model to cater to all existing variations. The model is to be implemented in a relational database with semantic data structures to facilitate multi-criteria searching and analysis.*

Introduction

In the era of information technology, public administration has been reformed from being a single system to being an integration of multiple systems. In other words, the traditional paper file system stored within an administrative unit has had to be replaced by a more comprehensive digital system integrating information from external units. In the past, the matching of individual records across different administrative units had to be done manually. Inconsistencies in land information arising from varying definitions, formats, languages, descriptions, and so on had to be resolved by individuals with experience. The advent of geo-information technology has led to calls for interoperability and the sharing of data to achieve a more effective overall government administration. A carefully designed system with the intelligence to resolve conflicts between numerous sub-systems has become necessary. In this regard, a system for matching addresses is our prime concern. Such a system is fundamental to the integration of all sorts of information on land. An address is the first essential description of any piece of land for whatever use: residential, commercial, industrial, recreational, or institutional. It is also a more permanent and legally accepted description as compared to other attributes such as use and ownership. In many aspects of public administration, an address is the only way to match various data sources.

As early as the 1970s, the DIME (Dual Independent Map Encoding) system was developed for the U.S. 1980 Census. It defined street and block relationships using a topological network of street segments, intersections and enclosing blocks (U.S. Bureau of the Census 1990). The system was later incorporated and refined in the Census Bureau's TIGER (Topologically Integrated Geographic Encoding and Referencing)/Line File for the 1990 Census, with the aim of reassigning the addresses to the

correct geographic block location. The TIGER/Line File has the important ability to assign geographic classification codes to the addresses of individual structures and to the names of buildings, and to catalog census-recognized geographic entities and their characteristics using a single relational database. Inconsistencies and errors that result when independently developed data files are used could thereby be prevented (Sobel 1990). Yet the two systems place too much emphasis on the rigid topological network of an address system that is merely two-dimensional. The heavy dependence of street networks to devise an address may result in ambiguity when a property sits at the corner of a block or at the intersection of two or more street segments. Properties located within a complex structure involving a third dimension of address description are not handled. Moreover, the coding of street segments and blocks is a unique but unfamiliar way of describing an address if implemented across address files of different sources. Recent developments have been geared toward properly defining the various road entities in designing a good logical database model. Examples include the model of hierarchical wayfinding for road networks (Car 1997) and the redefinition of terms such as spot, lane, pavement cross-section, segment, road, route, corridor, and region (Walls 2002). These are steps toward the formalizing of feature descriptions into a uniform standard. In consolidating the diverse address formats, Clayton and Bates (2002) devised an address-matching program that parses and cleanses a set of addresses, returning a set of validated addresses with an x,y coordinate.

Recent studies on address matching have mostly concentrated on linking the property onto a two-dimensional mapping system, thus making the implicit assumption that the relationship between built-up structures and property is one to one. There have been no discussions on more complex issues as property-property relation-



Figure 1. An example of address descriptions in different billing systems

ships, structure-structure relationships, and structure-property relationships. A complete addressing system should take all of these issues thoroughly into consideration, especially the issue of unnamed properties or structures within a structurally complex three-dimensional city. It is important to uniquely identify each property so that cross-departmental data searching based on an address is made feasible and convenient. For example, the Inland Revenue Department holds records of all properties owned by an individual for assessing the rental tax, whereas the Rating and Valuation Department needs these market values to assess the rental value of an individual property. The Census and Statistics Department keeps a very detailed breakdown of residential units down to the household for demographic surveys, whereas the Planning Department is more interested in the aggregation of these units for land-use planning. The property units for utility companies such as gas, telephone, and electricity, may share a similar definition, but this definition will be different from those used by the Buildings Authority, Lands Department, or Housing Authority. These are just a few of the many examples showing the need to share data based on a “commonly understood” addressing system. By this, we imply that the original address description may vary from one administrative unit to another, even though they all refer to the same property or piece of land. We should also point out that due to the high density of settlement in urban Hong Kong, a high “resolution” addressing system involving several attributes had been developed to pinpoint a unit. Such a scheme encourages the creation of many variations of the same address, making address matching more complicated. Before discussing how we can achieve what is “commonly understood,” let us first make a thorough study of the possible problems arising from the present addressing system. We will then present a scheme to structure land/property more systematically so that address descriptions can be made more semantic. It is noted that apart from textual address records, indexing through a map is also useful and is a better way of enhancing decision supports on spatial analysis.

Address Problems in Hong Kong—At a Glance

The major addressing issue discussed here is consistency. Figure 1 illustrates a case of different mailing address formats of the same property from various public/government organizations. In normal cases in the hierarchy of spatial units, an address is composed of a secondary planning unit zone (such as the New Territories, Kowloon, Hong Kong Island), a district (such as Shatin, Mongkok, Central), a building name, a street name and number, a floor or storey number, and, finally, a flat or apartment number or code. In most cases, a selection of essential items from the list is sufficient to match the concerned property with the various address formats, such as those shown in Figure 1. However, “what is essential” poses the biggest problem. Clearly, in this example, both the water and power bills record the most detailed descriptions of address. This includes the street name and number (9-11 Sha Tin Wai Road), the site development name (Greenfield Court), the building name/code (Tower or Block 2), and the storey(s) and flat belonging to the same property (20/F-21/F, Flat C or Duplex C, 20/F). Even this is not fully compatible when matching is performed, because of the different naming system used or the vague definition of “duplex.” Other items containing less detail also can be mailed to the same property, such as the omission of the 21/F in the gas bill. Nevertheless, automated matching between the rates/rent bill and the Treasury will create problems, because the former omits the name of the site development, while the latter leaves out the name of the street. The building code “Tower 2” and other descriptions are certainly not unique enough to link the two records.

The above example points to just a few problems in address matching. In fact, the difficulties of the Hong Kong addressing system are attributed to two major aspects: structural complexities and inconsistent descriptions. Structurally, on the condition that every land feature shall be given an address, what constitutes a unique feature or structure can be very difficult to say. In a structurally three-dimensional city, the feature can just be a park or open space on the topographic surface that may not involve any built-up structure. At the other extreme, it can be an extensively built-up structure extending vertically from below the topographic surface to non-uniform multiple storeys above the surface, and horizontally to adjoin neighboring structures. Address matching will be difficult if a feature is defined differently between administrative parties. For instance, a household means a whole flat with one unique flat number or code according to records on buildings, ratings, or water supplies; whereas a household might be further subdivided into rooms or just a bed inside a flat or even indicate a street sleeper according to census criteria. Similarly, a building is a legal built-up structure such as those declared to be a building by the Building Authority in a notice in the Gazette; whereas illegal structures such as on rooftops and balconies are also included in the charging of government rates and rents. Car parks, club houses, and the like belonging to but physically detached from the same residential unit may or may

not be included in one addressing system. Consequently, address matching of these features between one party and another may cause problems.

In another aspect, a most important description of the feature is its address. Here, a great deal of inconsistency can be found. Hong Kong does not possess a postal code system. The minimum partition size of the territory that is generally understood is the district or site development, such as Central or Queen Mary Hospital. A district is given a unique name, and all districts exhaustively cover the whole territory without any gaps and overlaps. The district boundary is a common but not compulsory element in addresses, and is used for many purposes, such as delineating voting regions and planning school place allocations. A site is a certain type of land use permitted for development by the government or private developers. This can be for recreational purposes like Ocean Park, industrial purposes like the Tai Po Industrial Estate, institutional purposes like a hospital, or residential purposes like Whampoa Garden. The site name is usually unique enough as to not require the district name in an address, but it is noted that a site may occur inside or across more than one district. Nevertheless the gazetted address can be without both district and site names. Before the developer proposes a site name, the Lands Department may have already identified an area with a unique lot number, despite its unpopularity with the public. The Rating and Valuation Department can also assign a street number to uniquely address its interested property. Like a site, the street may be so long that it crosses several districts and sites, making it difficult to construct a one-to-one relationship between these three elements. In terms of a building or structure, a small lot or land parcel, addressing is possible in the following ways: a) a lot number; b) a street name and number; and c) a district name, a site name, and a building name. Hence, under Section 2 of the Building Ordinance, it is also stated that there is no unique or simple rule in defining what a building is. A street name, street number, and/or lot number may be used as the address of a new building project. Some peculiar cases have given rise to greater complexity. For example, in recent years, there have been buildings with two different names, one for all single lower floors and another for the duplex apartments on the upper/top floors. Both the same or different names may be kept depending on the original records of the administrative unit.

The above discussion has only pointed out the varying address formats in the horizontal facet. Vertically, there is also no uniform address system for the whole territory. Disregarding what is generally understood as the first floor, second floor, third floor, and so on, the storeys or floors below the topographic surface can be called various names, such as upper basement, lower basement, upper ground, and lower ground. The different names would present no problem if meanings were consistent, such as upper basement always indicating the first floor below the surface. However, there is no common agreement on the semantics behind these names, as shown in Figure 2. A building constructed on a slope surface may adopt the same convention as a building on flat land. Similar inconsistencies also apply to

top floors. A more peculiar case is that some buildings omit the 13/F or 14/F because of certain beliefs about unlucky numbers. A more problematic issue is the bilingual address descriptions in Hong Kong. In some buildings, both the Chinese and English versions of 1/F, 2/F, etc., correspond to each other. In others, they are one floor away; for example, the Chinese version of G/F means the English version of 1/F, 1/F means 2/F, and so on. With these inconsistent descriptions, if a party wants to make certain comparisons or analyses based on the vertical facet, such as determining the average value of all apartments on the first five floors of a building, a great deal of human interpretation and editing are deemed necessary.

The previous paragraphs have exemplified some of the problems in our current addressing system. A more unified approach shall be adopted to facilitate information exchange through indexing an address. The model proposed in the coming sections will first address how structural complexities of land features can be more systematically defined, followed by a discussion of supplying an appropriate description to each component. It is hoped that most of the previously discussed problems that arise from or result in inconsistencies among various official address records can be avoided.

The Proposed Compartmentalized Model

The proposed model deals with the complexity of the Hong Kong addressing system and supplies a standard set of concepts and terminologies to identify a settlement unit in an urban environment. Concepts such as the secondary planning unit zone and district, which are often used in addresses, already have been standardized, and there is no need to address them in this paper. The purpose of this proposal is not to standardize the format of an address, but rather to supply the terms that can be used for sharing addresses among the various organizations. These terms can be mixed in various ways in an address, and we believe that

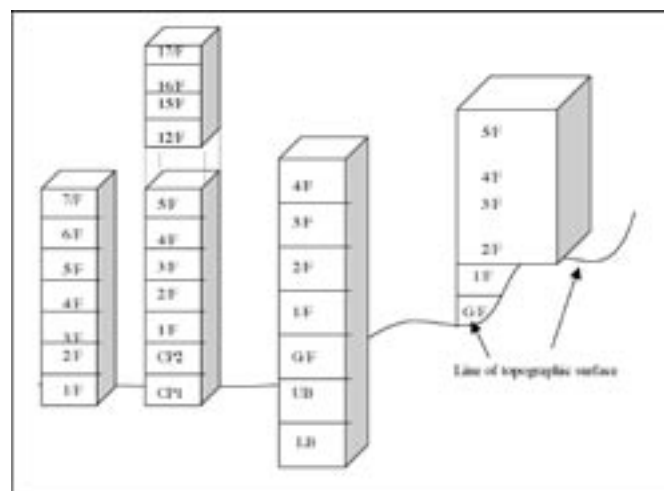


Figure 2. An illustration of the varying floor labels

they are sufficient to provide a unique address if they are mixed in the proper way.

The main features of the proposed model are the different units of settlement (the compartments) that form a hierarchy to describe settlements or building structures of various degrees of complexity. An address system often reflects the architectural characteristics of a building in which the unit is located. Although we have developed the present model for Hong Kong, we believe that it can be used in other urbanized areas as well. Other than some possible cultural differences in architecture, the building structures in Hong Kong are complex enough to encompass the building structures of other cities. It should also be noted that the purpose of the proposed model is not to re-create the structure in which an address is located. Instead, it is to be used to describe the location of a settlement unit within a complex structure.

Hierarchy of Compartments

To start with, the five compartments are identified as: “property unit,” “building block,” “building,” “development,” and “site.” The discussion that follows focuses on the standardization of the hierarchy between these compartments rather than on devising a formal and unambiguous definition for each compartment. We believe that users, in general, have an intuitive understanding of these compartments. Besides, the hierarchy so suggested has incorporated both the horizontal and vertical dimensions of all buildings or construction facets.

Property unit. A property unit (or, simply, unit) is a basic entity in the address database. It is a self-contained space or structure designed for a single commercial, industrial, institutional, or residential occupancy. Units are physically separated from each other. Examples of a unit include a room occupied by a single family, an apartment/flat, a house, a shop, an office, a school, a hospital, a clinic in a shopping mall, a department store, a garage, an open parking lot, or a deserted lot. It is noted that some units may have changed over time, meaning that the current partition of a property for renting to different households or businesses may have changed over time. This additional complexity of temporal inconsistency in addresses will not be discussed here, and it is assumed that all parties are interested in the most up-to-date data.

Building block. A building block (or, simply, block) is a range of contiguous floors, sharing very similar footprints and addressing systems. In most cases, a high-rise tower is equivalent to a block. This model allows the flexibility of subdividing a tower into multiple blocks so that they can be individually named, such as different names for the lower floors and upper floors. Most of the block footprints have been captured as a “building” in the large-scale 1:1000 topographic plans drawn up by the Survey and Mapping Office.

Building. A building is composed of one or several connected building blocks sharing a common outside wall (Figure 3). The different

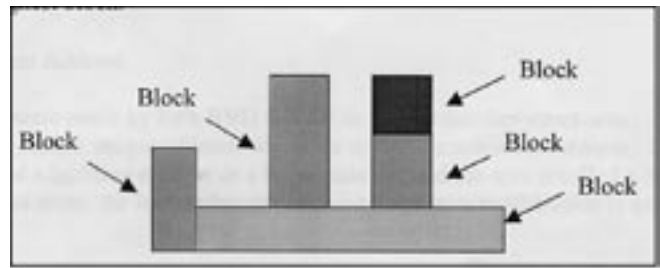


Figure 3. A sample building composed of blocks

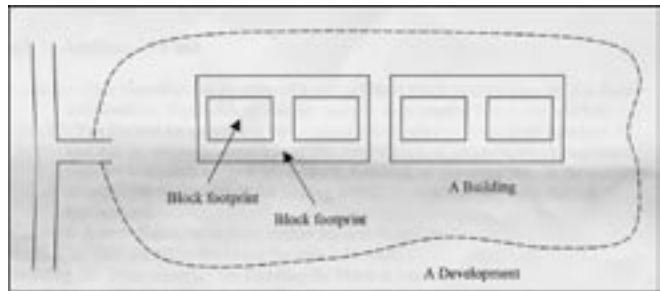


Figure 4. A diagrammatic representation of a block, building, and development

blocks may have dissimilar functional uses such as shopping malls or car parks for the lower floors, and residential units above the podium level. It is common for the blocks in a building to share the same external entrance. The footprint of the building is the union of the footprints of its blocks. A unit is contained entirely within a single building but could cross blocks (e.g., the apartment and car park belonging to Mr. Chan could cross different blocks). In general, the floors in a building are numbered sequentially from the lowest block to the highest block.

Development and site. A development is a group of buildings under a single management developed at about the same time (Figure 4). Buildings within a development are therefore relatively uniform in structure. A site is any contiguous area of interest as a single entity. It is as large as or larger than a development. Their delineation is like that of Tai Koo Shing (a site), which contains Tai Koo Shing Phases I, II, and III (phases of development). It is also noted that a development or site may contain buildings and blocks of varying land uses, and might be traversed by streets.

Address and Location Indicators

In the existing address system, the minimum address can simply be a number-street-district or a lot number-village-district. However, problems may arise if one address system records one street while another uses another street for the same feature. Compared with the compartment model of development, building and block names, street names and numbers are less preferred for the following three reasons:

1. They can address up to a single building block level only, but are not able to differentiate between different blocks in a building.

Block	Building	Development	Site	Comment
N/A	N/A	N/A	A	A site called A without structures.
A	A	—	Lot no.	A single block building called A. The block and the building share the same name.
A	A	—	Lot no.	A block without a name in building A. The building name propagates to the block.
—	—	—	Lot no.	A building without any unique name; e.g., latrine.
A	B	—	Lot no.	Block A in building B.
A	B	C	C	Block A in building B within development C. The site name propagates to development.
A	B	C	D	Block A in building B within development C at site D.

Table 1. Some of the combinations that could make up an address and the way name propagation works. N/A means not applicable whereas “—” means no name. Lot no. is optional.

Source	Block	Building	Development	Site	District
Water Supplies Dept.		2	Greenfield Court		Shatin
CLP Power Ltd.	2			Greenfield Court	Shatin
Towngas	2		Greenfield Court		Shatin
Ratings & Valuation		2			Shatin
Treasury		2		Greenfield Court	Shatin

Table 2. An example of how an address is recorded in the proposed compartment database.

2. In some large developments consisting of several buildings, street addresses are aggregated as, for example, “1-10, ABC Street.” In singling out a certain building, it makes no difference if only the building and development names are used while the street name is omitted.
3. Using the site, development, building, and block names as addresses can better clarify their structural relationships and enhance address-matching semantics.

With the help of the proposed model, we then can develop a neutral database for address sharing and interchange. In this neutral database, the proposed compartment terms will be used to identify an address. Although we have identified five compartments that could be used in an address, it is not always necessary to have a name for all of these compartments. When a search is performed on such a database, slight differences in the interpretation of an address could occur. For example, for a single-block building, the building name and the block name could become one, and the user might interpret a name in either way. In order to facilitate searches, we propose the following six rules for populating the address database:

1. A property unit must be associated with a piece of land (e.g., a site). In land cadastres, all sites are assigned a unique lot number. Being an address, the lot number is sufficient for a vacant site and for an undivided property in the third dimension. The usual practice is for any form of development on a site to be given a name by the private developer. If there is no more than one building on a site, its name will suffice to constitute a minimum address. That is, the building name bears the same address

identifier as the lot number, in which case the latter becomes optional.

2. A building name propagates to a block name, whereas a site name propagates to a development name. In other words, an unnamed block inherits the building name and so on. This is because block/building and development/site concepts are interchangeable when they refer to the same entity, such as a single block building and a single development site. Sharing the same name would facilitate searching.
3. The existence of a block implies the existence of a building, which in turn implies the existence of a development and thus of a site.
4. The existence of a site does not imply the necessary existence of a development. The site can be a vacant site waiting for development.
5. The existence of a building implies the existence of at least one block.
6. The existence of a development implies the existence of at least one building.

Let us illustrate the concepts with the varying address descriptions in Figure 1. In the database with the compartments identified, the operators may input the addresses differently according to their own understanding of the records at hand (Table 2). Some may interpret the structure labeled “2” as a block or as a building. A similar inconsistency would apply to the development and site unless there is an address with all four items labeled. Nevertheless, with the propagation rules that have been discussed, these varying address formats can be linked together to better identify the same property, compared with the original textual descriptions. One

obvious difficulty is with the Rating and Valuation records, which do not include the street name and number in the compartment database. This may be handled by adding one more column for street name and numbers, but for most cases the method is not recommended. This is because: a) this will add additional complexity to the database, and b) street names and numbers may not be unique among different address sources.

To identify the location of a property unit within a block, the number-flat-floor combination is sufficient. In order to enhance compatibility and consistency in floor labeling within the same block, an alias can be inserted. For example, in some cases, 1/F equals to G/F, podium level equals to upper ground level. A common table listing the standardized floor labels and other different options is required. There should be rules governing the continuation of floor numbers across different blocks of the same building or separate floor labeling for discrete but contiguous blocks.

The minimum address and/or other address qualifiers make up the textual form of the address, which are good enough for postal purposes. However, for spatial display and analysis, it is desirable to have a more concise and unique geo-referencing code for the compartments. We propose the use of footprint centroids for such applications. In this way, a centroid can be associated with each unit and for each compartment. If two vertically aligned compartments share the same centroid, one of them could be slightly shifted against the other to avoid collision. Such a centroid also could serve as a unique geo-referencing code across organizations. It is worth noting that such a mapping index in the address database is extremely useful in a multiplicity of pattern analysis, compared to a mere individual unit or property search in the textual address database.

Conclusion

Based on the Hong Kong setting, an addressing model emphasizing land feature compartments is proposed. Yet, the same principle might be applied universally to other cities. Features are first identified structurally and spatially in a hierarchy, and each name in the address reflects the hierarchical level it belongs to. The suggested compartments of “block,” “building,” “development,” and “site” can well apply to other urban settings. A different set of notations and number of levels might be adopted, but the most important criterion is the nesting of one compartment into another to build up an address. In places with only simple building structures for one identical use, the block compartment may be eliminated. On the other hand, if a piece of land and its development are concerned with multiple levels of ownerships, additional compartments beyond “site” can be defined. The main principle of the model is to allow propagation of an address or name from one level to another, so that cross-matching of information at the same level is enabled. This approach also helps relating the textual address description to its spatial representation in a digital mapping or geo-information system. On the other hand, with the exception of the bilingual characteristics of floor labeling that is quite unique in Hong Kong, the suggestion of an alias label (in relation to the topographic surface) for aligning floor descriptions

across different buildings also might be considered in other urban settings with similar problems.

In summary, this paper points out the design of a number of concepts related to the making of a meaningful address system. The discussion has focused on theoretical considerations rather than on the detailed design for a database. The main objective of this compartmentalized addressing model is to facilitate the interchange of property or land-related data among user databases. There is no intention to override or disapprove of any existing official address system. Instead, in this new era that calls for data sharing and administrative efficiency, a neutral system integrating a wide variety of formats and concepts is important.

Acknowledgements

Many of the concepts and ideas presented in this paper were extracted from a consultancy project (1999–2000) with the Rating and Valuation Department, the Hong Kong SAR Government, in which the authors were then the GIS consultants from the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University.

About the Author

Dr. Lilian Pun is currently Assistant Professor in the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University. Prof. Y.C. Lee is Professor in the Department of Geodesy and Geomatics Engineering, University of New Brunswick. Research interests of both authors are in the areas of GIS, data modeling, and, in particular, urban applications.

Corresponding Address:
Department of Land Surveying and Geoinformatics,
The Hong Kong Polytechnic University
Hungghom, Kowloon,
Hong Kong
Fax: (852) 2330 2994
E-mail: lspun@polyu.edu.hk

Authors Note

It is with deep regret that the co-author, Professor Y.C. Lee has passed away in June 2004. Professor Lee was a great researcher and contributor to the discipline of Geo-Information Technology.

References

- Car, A. 1997. Hierarchical spatial reasoning: Theoretical consideration and its application to modeling wayfinding. Department of Geoinformation, Technical University Vienna, Austria.
- Cassettari, S. 1993. Introduction to integrated geo-information management. London, Chapman & Hall.
- Clayton, B., and C. Bates. 2002. Building an effective address-matching program and some practical applications. Highlighted papers from the August 11–13 Street Smart and Address Savvy Conference, Portland, OR, August 2002, <http://www.urisa.org/>.
- Sobel, J. 1990. Principal components of the Census Bureau's TIGER File. In D. J. Peuquet and D.F. Marble, eds., *Introductory readings in geographic information systems*. London, Taylor & Francis: 112-119.
- U.S. Bureau of the Census. 1990. Technical description of the DIME system. In D. J. Peuquet and D. F. Marble, eds., *Introductory readings in geographic information systems*. London, Taylor & Francis.
- Walls, M. D. 2002. Tools for modeling complexity—The street network example. Highlighted papers from the August 11–13 Street Smart and Address Savvy Conference, Portland, OR, August 2002, <http://www.urisa.org/>.