# An Orthodox Approach towards Automated Bilingual Name Placement in Large Scale Maps

**Lilian S.C. Pun-Cheng**

Department of Land Surveying and Geoinformatics
The Hong Kong Polytechnic University
Hunghom, Kowloon, Hong Kong
email: lspun@polyu.edu.hk

## Abstract

A name or label is an essential and important component of a map. Yet, its placement has always been considered as a most labour-intensive process in manual cartography. In Hong Kong, the process is further complicated by the presence of dual languages (English and Chinese) together. Conversion of 1:1000 analogue map sheets to digital data have been completed but text placement is still performed in an interactive mode. This paper provides some thoughts on establishing effective and appropriate algorithms for the full automation of name or label placement in both languages. Tested samples are drawn from the 1:1000 monochromatic digital topographic map sheets, with the Arc/Info coverages established by the Land Information Centre of the Lands Department. The project aims not only at increasing speed, but also at searching optimum map location, maintaining logical hierarchy and consistency.

## 1 Background

A name or label is made up of letters (in English), characters (in Chinese) and maybe with numbers. Automation as described in this paper applies to all these components that constitute an entire name. For consistency, the term 'name placement' which is meant to be the same as 'text placement', 'lettering', 'annotation' and 'labelling' is used throughout the rest of the paper. Names and labels are essential and important components of a map. It not only tells us the nominal characteristics of the geographical feature represented, but also gives an implicit understanding of the linear or areal extent and orientation of a map feature as well as its relative size or importance. Labels or names associated with different map features can be treated as attributes, useful for data retrieval or spatial analytical operations if organised systematically. However, name placement may be considered as a most difficult and labour-intensive process in manual cartography. Even with the advent of computer technology in most cartographic processes, automated name placement is considered by far the most difficult to handle in view of the complexity of rules and topographic features that are involved. Several automated placement programs have been developed but most of their results are still far from satisfactory (Buttenfield and McMaster, 1991). For example, procedural knowledge has been used to guide the placement of typography on maps. Several successful implementations of using rule-based approaches to place names are associated with point symbols (Hirsch, 1982, Jones, 1989) or areal symbols (Pinto and Freeman, 1996) only. Little is talked about the complexity of associating names with all point, linear and areal features which are common on topographic maps. Consequently, some automating techniques have been noted for overlapping text, upside down text and text placed at awkward angles.

In Hong Kong, the conversion of 1:1000 analogue map sheets to digital form has been completed. But name placement (including English words and numbers) is still performed in an interactive mode. That is for instance, if building blocks are to be labelled, the central position of each block or polygon has to be determined or adjusted manually by the user and the name is placed accordingly. Such is in fact a very time-consuming, inconsistent and inflexible method. It is estimated that at least 8 man-hours are needed to complete labelling a full A0 size 1:1000 map sheet. Also at changing scales, repeating determination of text position, font and dimension for different features is required. The process will further be complicated by the need and future plan to incorporate Chinese place names (in form of discrete Chinese characters) as well on these

Journal of Geospatial Engineering, Vol. 2, .No. 2, pp. 37-43.

general-purpose maps. Methods of how to put these two very different languages together have recently been a concern of many software vendors in the region. Notwithstanding these, in view of the voluminous task of managing, handling and editing land information, there arises a need to automate the name placement procedure, so that it can couple with other GIS functionality for more optimal use of the technology.

## 2 The Hong Kong Digital Data Set

A project was then initiated to investigate the dual language (English and Chinese) name placement on digital maps of Hong Kong and to establish effective and appropriate algorithms for the full automation of the task. An orthodox approach is taken in that existing analogue maps are reviewed to formulate basic conventional rules for name placement. Amongst the numerous textual information that appears on maps, this paper focuses on the discussion of placing building and roads names on large-scale plans. It is because these two sets of information are relatively more important to most urban users in the region compared to point features labelling of for instance spot heights and lamp posts. Besides, special attention should be paid to their dual language requirements in labelling. Before that, it is necessary to examine the existing map data structure, so that algorithms developed will, as much as possible, not involve drastic restructuring of the data set.

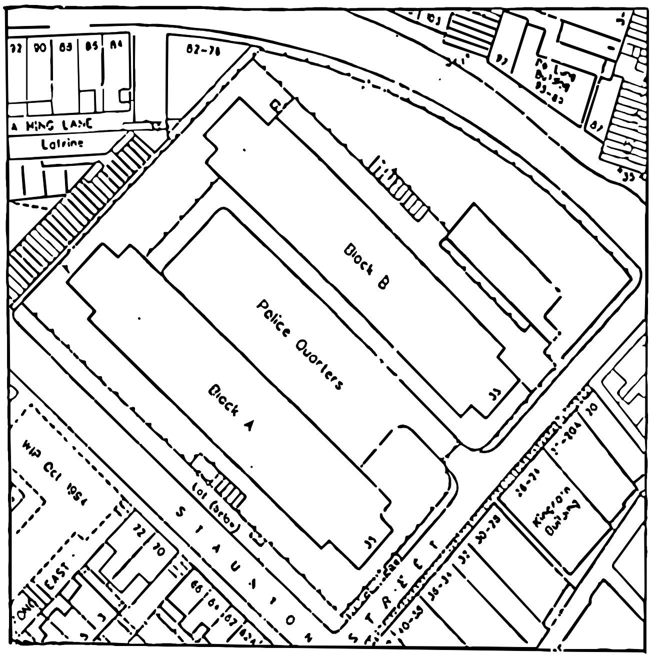


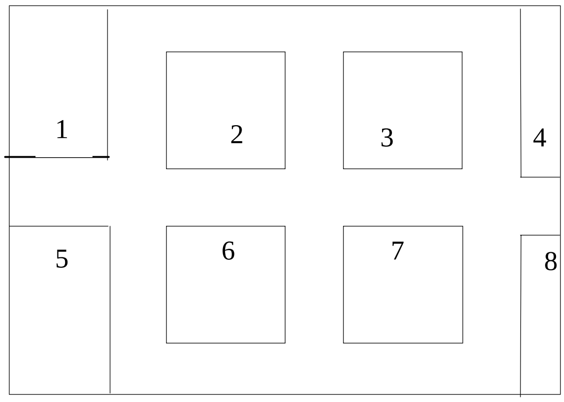

**Figure 1. Textual information of 1:1000 maps, Hong Kong.**

Textual information of existing 1:1000 plans mainly falls into three categories (Figure 1): names of area features of buildings and roads; house numbers of buildings along the side bordering the pavement; and labelling of point features like spot heights and lamp posts. By area, side or point, we refer to the geometry being considered when placing a name or label. If this conforms to the geometry of the feature being represented digitally, algorithms for name placement may be derived and implemented fairly easily within the GIS or any CAD annotation environment. However, the data structure now employed clearly deviates from this assumption. For buildings and all point features, these are represented in polygon and point geometry correspondingly and so pose least problems. The greatest difficulty lies with naming roads. While a road name should refer to the whole stretch of road, at present a road is being represented in discrete line segments (Figure 2) and so is impossible to extract it as one whole polygon. To resolve the problem, discrete line segments bordering a road are joined together with, if necessary, the addition of dark (invisible on graphics) line segments connecting adjacent road blocks (Figure 3). Then from the derived road polygon, a dark road centreline is generated for guiding name placement.

## 3 English and Chinese Characters

The characteristics of different English typographic styles and forms with their selection criteria have been described in many texts (Keates, 1989 and Robinson et.al., 1995). The following summarises the major differences with their Chinese counterparts only in modern or conventional ways of lettering (Figure 4):

a) while English letters may vary in size with the ascenders and descenders, a Chinese character may all be preserved in the same size and probably fit into a square of predetermined size;
b) naming a feature in English has to take into account of the spacing of both its words and letters (a name consists of one or more words while a word is made up of letters), whereas the corresponding Chinese name only considers spacing of a few characters (which are English equivalence of words); and
c) the conventional way of English reading starts from left to right while both directions may be adopted for reading Chinese names.




| *Road-id* | *Length (m)* | *Name* | *Remark* |
|---|---|---|---|
| 1 | 3.8 | Nathan Road | road margin |
| 2 | 12 | Nathan Road | road margin |
| 3 | 12 | Nathan Road | road margin |
| 4 | 2.9 | Nathan Road | road margin under structure |
| 5 | 3.8 | Nathan Road | road margin |
| 6 | 12 | Nathan Road | road margin |
| 7 | 12 | Nathan Road | road margin |
| 8 | 2.9 | Nathan Road | road margin under structure |

**Figure 2. Spatial representation of roads and associated database structure.**

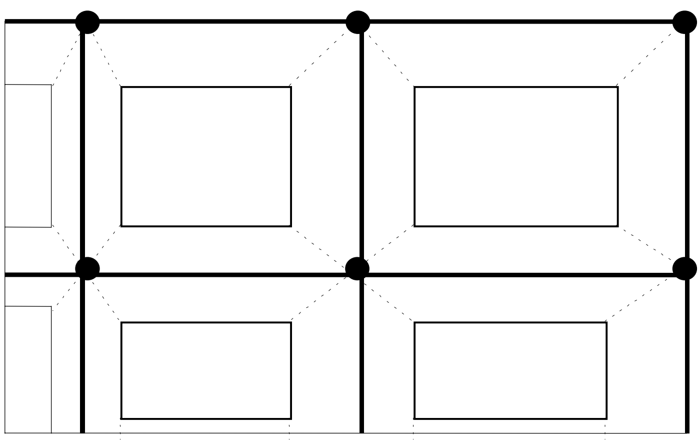


**Figure 3. Dark line segments (dotted line) and derived centerlines (thick solid line) and junction nodes.**

For annotating features in both languages, it is advisable to separate the two languages altogether, meaning not to alternate them with words or characters. The custom is to place the Chinese version on top of the English one for roughly symmetrical polygons; and to put them side by side for elongated polygons like roads. For the sake of consistency, especially for those map readers who are literately bilingual, naming in Chinese also starts from the left as that in English. Lastly, in case space is, after all means of considerations as discussed in later sections, still not sufficient for placing both languages, a rather subjective compromise has to be made in that the Chinese name would give way to the more popular use of English.

士 他 花 利 街

**STAVELEY STREET**

**Figure 4. Chinese characters and English**

## 4 Name Placement Algorithm for Buildings

The procedures that are discussed below are all implemented in the Arc/Info system, with the algorithms for labelling buildings and roads names written in AML (the Arc Macro Language). They are tested with a 1:1000 map sheet in the Central Business District, covering the busiest part of Hong Kong and where a dense network of roads and constructions is found. In this section, the algorithm for placement of building names is presented and the algorithm for placement of roads will be presented in next section.

### 4.1 Preliminary Considerations

As 1:1000 is the scale used in labelling buildings, no classification of buildings is performed. It is therefore essential to label all buildings, regardless of their ground areas, with the same font size, indicating their equal order of importance. In view of the numerous small-sized buildings, a font size of 1.6mm, which is close to the threshold of perception, has been selected. Ideally, the name should be wholly placed inside the building polygon. However this may pose difficulty with long names and small polygons. Hence a database consisting of original long names, possible short forms or abbreviations and separation into a few lines is prepared. This almost is the exclusive requirement for English instead of Chinese names. It is also considered that as most buildings are symmetrical in shape, a calculated central point is sufficient to guide placing the name in whatever desirable direction. Though infrequently occurred, for buildings of asymmetrical shapes, such a system-generated central point may occur outside the polygon and so has to be moved inside interactively (Figure 5).

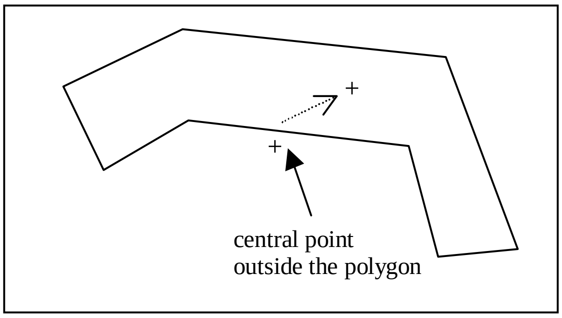



**Figure 5. Interactive edit of the central label point.**

### 4.2 The Approach

With all the preparatory information ready – a database of long and short names, spatial information (geographic co-ordinates) that forms the building polygon, polygon identifier and central label point, the program starts to search and determine the best position for name placement. The text length (in dual languages) in full is first compared with a horizontal distance centred at the label point. The horizontal direction is set as parallel to, in preferred sequence, the northings (map horizontal grid lines) or the longest axis of the polygon. In addition, three tolerance circles – one circle at the centre to test the text height of the dual language and two circles at the ends to test if the English text will cut or allow sufficient space apart from the polygon edges (Figure 6). In case these are not satisfied, spacing between Chinese characters and/or English words and letters have to be adjusted for repeated testing until at a dimension that differentiation between individual character, word or letter is difficult. If all fail, the same procedure will be tried for short names and, if necessary, splitting into several lines. For the latter case, the dimension of the orthogonal direction (to that of the horizontal direction) to cater for several lines has to be tested. Besides the number of words per line and if the names are separated sensibly are also important considerations. Surely this requires more complicated and longer testing than before. In summary, there are two main principles for placing building names. First is to place the name in fewer lines (preferably one) and horizontal to the map orientation for faster and simpler processing. Second, with a view of being more informative, the full name is preferred to its abbreviated form.

### 4.3 Sample Results

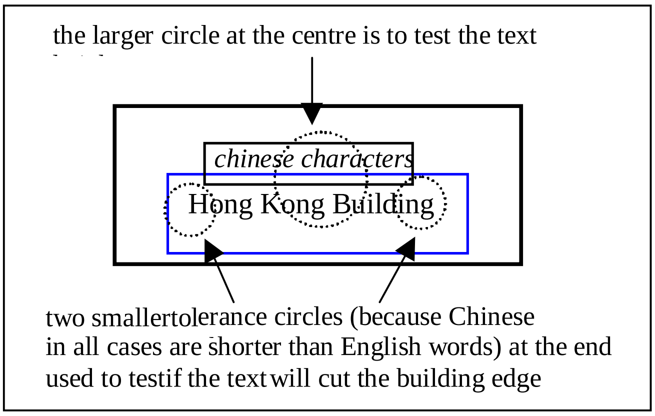



**Figure 6. Testing dimensions with tolerance circles.**

The tested area consists of a total of 106 buildings which require name placement. Satisfactory results account for about 60% while another 20% may be improved by interactive editing. These generally arise from buildings of extremely irregular and awkward shapes for the program to handle (Figure 7). The rest of the building polygons are just impossible to wholly inscribe a name owing to either their being too small or having too long names. In fact, conventional manual or interactive CAD-assisted name placement meets with the same problems and these buildings are also left unnamed in analogue map sheets.

## 5 Name Placement Algorithm for Roads

In previous section, the algorithm for placement of building names is presented. The algorithm for placement of roads will be presented in this section.

### 5.1 Preliminary Considerations

Unlike buildings, roads even at this large scale may be classified into three categories: main roads with more than one lane, secondary road with only one lane and minor roads of walkways. These are already varied by nature in length and width. So a certain font and size may be consistently applied to each class. In other words, appropriate road naming can reflect to map users about the road class. With some exceptions of minor roads, road names do not normally require short forms or abbreviations owing to placement constraints. However name placement of roads do suffer from two main problems.

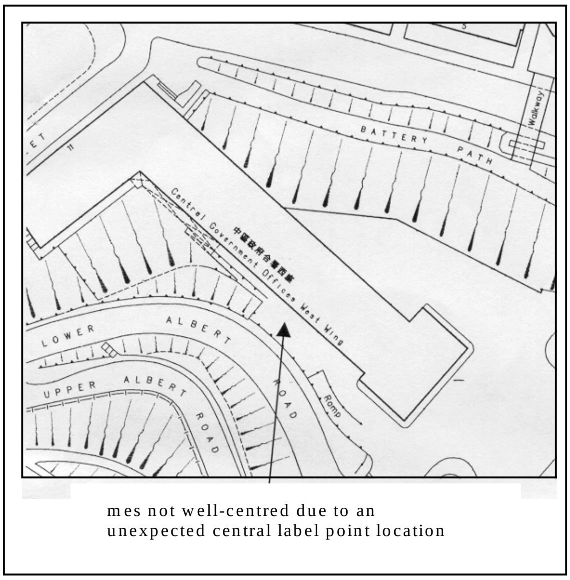



Figure **7. An example of awkard-shaped building that requires further interactive editing.**

One is concerned with overlapping texts at the junctions, particularly for two intersecting minor roads with long names. The other is deviation from the road alignment. It is important that the centreline created (which does not appear on the original map) for guiding the placement should align with the original road curvature. Also for long roads, it is necessary to repeat labelling at appropriate intervals.

### 5.2 The Approach

The font size and spacing has already been predetermined for each class of road, being 2.4mm, 2mm and 1.6mm for individual character or letter for main, secondary and minor roads respectively. Also a spatial database of road centreline information such as length, number of segments and junction nodes that make up a road is available. A road segment is defined as a stretch of road (whole or part) from one junction to another. The algorithm differentiates between the different classes of roads as well as between single-segment roads and multi-segments roads. It handles name placement from minor roads and those with single segment first to major roads with multiple segments, in the order of increasing flexibility. For single-segment roads which are mostly the minor roads and some secondary roads, both the English and Chinese names are placed only once. The name is adjusted to appear in the middle part of the segment with a predetermined offset distance from the two ends. Further adjustments on text spacing are catered for very short segments, and in the event of extreme cases, the Chinese version or both have to be removed.

For multi-segments roads, the text length in dual languages have to be compared with the total centreline length to determine the necessity of repeated labelling. Repeated occurrences of road names have to be separated by a reasonable length, and by default here one segment length. The offset distances from the two ends of the entire stretch of road have also to be decided. On the other hand, to avoid overlapping texts at the junctions, a tolerance circle centred at the junction node is used as constraint. This means that the texts are not accepted to be placed inside this circle. Now, labelling starts from the segment closest to the mid-length of the centreline. The ideal case is that this middle segment is long enough for both languages. Else, either the English name or the Chinese characters have to be placed in the adjacent segment. In case the middle segment is too short for either language, it is omitted and adjacent segments are used. This in fact seldom occurs for main roads which usually border large street blocks. Once a segment is named, the adjacent one is left unnamed. Alternations of ‘placement’ and ‘no placement’ and of languages are performed using the same reasoning as the middle segment. The process continues until meeting the offset distance from the two ends of the entire road or of the map margin. If, under exceptional circumstances, all segments are too short for the name in either language, part of the name will be allowed to extend into the tolerance circle, i.e. the junction area, provided that it does not interfere with names of intersecting roads.

### 5.3 Sample Results

Except for very short roads which may not accommodate the length of a road name in two languages, most name placement of roads are satisfactory. The success rate is as high as 80% without further interactive editing. However running time can be longer for very winding roads and those with numerous segments and junctions. Though occurring only in very few cases, there are still problems of (Figure 8) overlapping texts at the

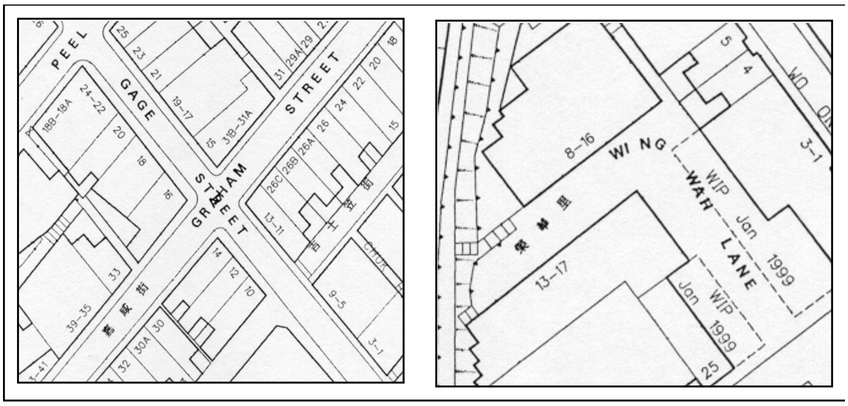



**Figure 8. Examples of unsatisfactory results in naming roads.**

junctions as well as unsatisfactory results for L-shaped minor roads. It is recommended that if found necessary in other tested areas, names of roads can also be abbreviated but of course at the expense of a more time-consuming database construction and program execution.

## 6 Conclusion

This paper has presented the procedure of automatically implementing some orthodox theories on name. For users of large scale plans as the 1:1000 to 1:5000 scale range, accurate textual information is extremely important for orientation, location query, everyday routine operation and so on. It is found that the addition of another language does not pose too much difficulty or complexity to the task. One promising thing about automation is speed. What takes a day's work in naming only the buildings and roads of one A0 sized map sheet can now be fulfilled in about 15 minutes. There are certainly more benefits in automation like reliability, efficiency that need not be elaborated here. Nevertheless, one very key issue to the success of name placement algorithm is not on its hardware or software requirements, but the compatibility of the underlying spatial data structure. The way that we perceive a cartographic feature (as a point, a line or an area) and its corresponding spatial representation as a whole object or entity do facilitate the automation process, from the conceptual design of lettering rules, to formal specification in the algorithm and ultimately to final implementation. That part on cartographic modelling or land data modelling is beyond our scope of study here, but a good model is extremely important for numerous and cross-disciplinary applications, among which automatic name placement of the mapping profession is one example. This too is a growing emphasis on recent developments of spatial data and information system.

## Acknowledgement

The work described in this paper was substantially supported by a grant from the Hong Kong Polytechnic University (Project No. G-S118).

## References

Ahn, J. & Freeman, H., 1984. AUTONAP – an expert system for automatic name placement. *Proceedings International Symposium on Spatial Data Handling*, pp. 544-569.

Buttenfield, B.P. and McMaster, R.B., 1991. *Map Generalization: Making Rules for Knowledge Representation.* Harlow, Essex, England: Longman Scientific & Technical; New York, N.Y.: Wiley.

Dijk, S., Kreveld, M., Strijk, T. and Wolff, A., 1999. Towards and evaluation of quality for label placement methods. *19th International Cartographic Conference Proceedings* 1: 905-913, Ottawa, Canada.

Hirsch, S.A., 1982. An algorithm for automatic name placement around point data. *The American Cartographer*, 9(1): 5-17.

Jones, C.B., 1989. Cartographic name placement with Prolog. *Industrial Electronic and Electrical Engineering Society Computer Graphics and Applications*, 9: 36-47.

Keates, J.S., 1989. *Cartographic Design and Production*, 2nd edition. Longman Scientific and Technical, Essex, England.

Leskinen, T., 1999. National place name register integrated with cartographic names register for multiple scales and products. *19th International Cartographic Conference Proceedings*, Ottawa, Canada. 1: 885-892.

Pinto, I. and Freeman, H., 1996. The feedback approach to cartographic areal text placement. In Rosenfeld, A., Perner, P. & Wang, P., editors, *Advances in Structural and Syntactical Pattern Recognition*, Springer, New York. 341-350.

Robinson, A., Morrison, J., Muehrcke, P., Kimerling, A. and Guptill, S., 1995. *Elements of Cartography*, 6th edition. John Wiley & Sons, Canada.