CARTOGRAPHIC DIGITISING - TECHNICAL TRENDS AND ECONOMIC FACTORS

P. A. Woodsford

Laser-Scan Laboratories Ltd.,
Science Park, Milton Road
Cambridge CB4 4FY

ABSTRACT

Automatic digitising of map data has made substantial progress, with improved techniques for both raster and vector scanning. Further progress can be anticipated, but effective and economic mass digitising remains a bottleneck. The paper discusses a classification of the wide range of requirements for digital map data. The digitising complexity of these requirements, together with issues of accuracy and quality both of input documents and of required digital output dramatically affect the applicability and the economics of techniques for automatic digitising.

INTRODUCTION

The more senior British members of the audience may well recollect a very worthy BBC radio programme of years gone by called "The Brains Trust". One of the leading members of the Brains Trust was the late C E M Joad, who invariably prefaced his answer to each and every question by the famous words - "Well, it depends what you mean by.....". This response is particularly appropriate to the question "What will it cost to have these maps digitised?" This paper is an attempt to map out the various issues involved, against the background of rapidly changing technology, and to relate these to management and cost considerations. As such it has to be a taxonomy of both the requirements for digitising and the techniques available for digitising - a daunting task in a short paper, but one that can be undertaken in the context of Auto Carto London, where so much of the evolving technology and the motivating requirements can be studied both in the papers and the exhibition.

The problem with taxonomies of multidimensional phenomena is determining which is the key variable to use for the primary classification - how to skin this particular proliferating genus of cats. In this paper I will seek to order the phenomenon of the capture of cartographic information into computer systems by analysing primarily the range of requirements, using this classification as a framework for further considerations of technology and economics. This is, I believe, the most fruitful approach.

The data capture bottleneck is still with us, despite the advances in technology, although the problems, and the associated costs, are being significantly re-defined as the requirements are refined and better understood.

REQUIREMENTS FOR MAP DIGITISING

Maps as pictures

Rapid advances are currently underway in the enabling technology for handling and distributing maps as pictures, both in video form and in computerised systems giving scope for mixing map information with other information held in databases. One attraction of these applications is that the digitising requirement is particularly simple since basic raster data can be used. Digital cameras will suffice for low resolution requirements and the needs of the facsimile market are leading to the availability of cheap, medium resolution scanners.

Certain cartographic publication tasks can be automated by computerised graphics arts techniques although costs are still high and the benefits have to be realised entirely within the cartographic publication activity, there being no potential to create map databases or derived products such as terrain models.

Map backgrounds

The major users of large scale mapping are utilities (gas, water, telecommunications etc) who require a map background as the spatial reference for their plant records. Traditionally this has been accomplished by drafting plant records as an overlay on a map background. Analogous requirements exist in command and control display applications with the extra complexity of a time variable.

In computerised systems, map backgrounds can be generated by projection displays or by use of raster map data on raster displays. Raster map data is easily and quickly captured, but data storage requirements are high. It is difficult to build-in structure or feature coding and interrogation on the display is by position only. Care is necessary to compensate for distortion if, for example, paper documents are scanned, and coordinate transformations are computationally heavy.

The speed and low cost of creating raster map backgrounds may lead to their use as an interim solution even though many of the benefits of map information management systems are precluded.

Map Data for Display and Simple Interrogation

Most existing computer systems use map data in a vector form and the dominant and rapidly escalating requirement for digital map data at large scales in countries such as the U.K. is for this form of data. The requirement, by comparison with the basic mapping available from the national mapping organisation can be characterised as follows:-

(i) area cover. It is a strong requirement that substantial areas be covered in short periods of time. The perceived benefits can only come from the computerisation of whole areas.

(ii) partial capture. Not all the detail on the base maps is required, a typical selection being building outlines, property numbers/names, road outlines, road names, walls and fences and "miscellaneous linework".

(iii) simple feature coding. The data needs to be classified into a small number of layers, typically 6-10.

(iv) planimetric accuracy. Data needs to be captured and held to an accuracy criterion of within a half line width at source. In practice there may be some relaxation of this to cut costs, but this is not encouraged (and should not be).

(v) appropriate presentation. Within the requirement for planimetric accuracy, cartographic presentation standards can be quite relaxed. There is no requirement in this application for the fine positioning of names, or niceties of alignments for cosmetic purposes, particularly if these result in significantly higher data capture costs.

(vi) economy of representation. It is important that data volumes are kept near a minimum, and consequently that features in the data are represented by the optimum number of coordinates. This is particularly so when the lifetime costs of holding, using and maintaining the data are considered.

This requirement as it is currently evolving has been spelt out in some detail because of its market importance and because in certain fundamental respects it is at variance with the next requirement to review.

Map Data to Support Cartographic Products

Map Production agencies can support certain products by using "maps as pictures" techniques. However many requirements necessitate the creation of databanks (collections of unstructured datasets) if not fully structured databases, but the additional requirements of cartographic presentation lead to a different set of characteristics:-

(i) slow, progressive area cover

(ii) full capture

(iii) complex feature coding

(iv) planimetric accuracy essential

(v) heavy emphasis on a cartographically acceptable presentation.

Some of the conflicts between these characteristics and the preceding requirements are obvious; others have very important repercussions on the applicability of automatic digitising and hence on cost. These will be expanded later.

Map Data in Structured Form

The more ambitious aims of Geographic Information Systems present a requirement for map data in structured form, and a clear consensus is emerging that a link and node structure is the master pattern. Production experience of creating (and, more importantly, of using and maintaining) large amounts of map data in this form is not yet available and any special repercussions on digitising are not yet clear, although it is clear that the requirement for structured data increases the need for automatic, or semi-automatic, digitising.

Map Data for Terrain Modelling

This important requirement is used as an exemplar of requirements for non-cartographic purposes. It poses some requirements on map data capture in a particularly stringent manner, namely:-

(i) the need to reconcile data across sheet boundaries (this is present in all requirements to create continous digital maps but is particularly acute in terrain modelling).

(ii) the need to use appropriate source material, augmented if necessary. For high quality terrain models, source augmentation is more costly than digitising, especially if sources are defective in any way. The mathematical exercise of modelling exposes source mapping to a very thorough self-consistency check.

This excursion into a particular non-cartographic requirement is an example of the fact that digitising often involves much more than the conversion of existing graphic sources into digital files.

TECHNOLOGY AND QUALITY ISSUES

Manual Digitising

Most digitising is still done on manual digitising tables, and the advent of cheap microcomputers has considerably enhanced this approach since quite sophisticated checking and formatting software is readily available. Attempts have been made to increase the throughput of manual digitisers by adding a form of local scanning, but any successes using this approach have not been widely reported.

Manual digitising involves low capital cost and has a very fundamental advantage if source documents are of poor quality or in need of human interpretation. Slow speed and the difficulty of maintaining accuracy of output remain as fundamental disadvantages. Software can provide checks on syntax but not on content (i.e. accuracy of coordinates or of coding). Attempts to perform volume digitising by "cottage industries," unless very carefully managed, result in significant quality control costs.

Automatic Scanning

Most automatic digitising systems are based on the use of raster scanners, with a subsequent raster-to-vector conversion stage (usually software, but increasingly implemented in hardware), followed by editing and tagging cycles until clean data is produced. Scanning and raster-to-vector conversion times are impressively quick and progress is being made on symbol and character recognition.

The quality of the source document has a key effect on the applicability of automatic scanning. Blemishes or noise can throw the most sophisticated algorithms and create clean-up problems for the edit phases. Unless source documents are of a very high quality effort therefore has to be expended to prepare the source document graphically before scanning or to clean up the scanned data prior to vectorisation. In some cases the latter process can be substantially achieved automatically. A further point to note about automatic scanning is that the whole source document is scanned and processed, even if only a selected amount of information is actually required. A related issue is that it is difficult to apply specialised algorithms to particular classes of features (since to do so such features have to be recognised first). This makes it difficult to achieve optimal representations of say building outlines on an urban plan, and in consequence it is generally found that data volumes from vectorised raster data are greater than is desirable, with resulting increased costs in using and maintaining the data.

Hardware for automatic scanning is simple and costs are rapidly falling, since developments are underwritten by a market much wider than cartography. Software and procedural improvements continue, including separation of data into simple layers by differing physical characteristics (line weight, line style, colour) and the separation of text from linework, with subsequent text recognition.

As with all automatic processes it is essential in evaluating the cost effectiveness of automatic digitising to assess and measure the "bottom of the line" cost of the whole process, including all phases required to produce clean, accepted data.

Interactive Scanning

The alternative approach to automatic scanning is interactive scanning, or as it has somewhat misleadingly been termed historically, line following. The hardware required for this is a programmable scanner that can access the source in arbitrary patterns, and a means for an operator to follow and supervise this process. One particular system using this approach is the Laser-Scan Lasertrak.

Interactive scanning allows selective data capture, both in the sense of capturing only data that is required and in the sense of selecting the algorithm to be applied to a particular type of feature. This results in optimal, or near-optimal, feature representation and consequently near optimal data volumes. Furthermore feature coding takes place in parallel with scanning, although this advantage is reduced if feature coding schemes are complex. Because the interactive operator can resolve difficulties as they occur, and because high scanning resolutions can be used in local areas, interactive scanning can be less susceptible to the effects of defects or noise in the source document than automatic scanning. For example it is quite possible to effectively digitise data from a U.K. Ordnance Survey SUSI (Supply of Unpublished Survey Information) document on the Lasertrak.

Interactive scanning has the disadvantage of involving the operator with high cost capital equipment and an adequate cost assessment can only be made on a case by case basis. Developments continue aimed primarily at exploiting the flexibility of this approach and at optimising the production of structured (link and node) data.

The Applicability of Automatic Digitising

The applicability of automatic digitising, either automatic scanning or interactive scanning, depends critically on three parameters of the task in hand:-

(i) the quality of the source documents. The physical quality of the graphics may not be adequate for scanning or the content may be incorrect or inadequate to sustain the task. One of the fundamentals of computerising anything is that famous maxim "Garbage In, Garbage Out". To defeat this is expensive and likely to remain so.

(ii) the complexity of the coding required. If feature classes can be physically distinguished, either as separate overlays or different line styles, then automatic processes can cope effectively. If the number of feature codes is small and the classification is simple and unambiguous, efficiency of data capture is not seriously impaired. If the feature coding scheme is complex and "judgemental" the costs of feature coding will outweigh those of data capture, particularly when the costs of checking and quality control are included, as they must be.

(iii) the quality specification for the output. Costs rocket if this specification contains, explicitly or by implication, a subjective criterion such as "cartographic acceptability". Digitising is analogous to tracing not to fair drawing, if drafting analogies may be used. Automatic digitising is simply quicker tracing. It may, at a cost, incorporate elements of the fair drawing process and it is very desirable in many cases that it should do so. However the cost effectiveness of doing so can only be established if these additional elements of the task can be defined and quantified and so expressed algorithmically and checked reasonably.

SUMMARY FOR MANAGEMENT

Insofar as progress in Auto Carto still depends, in part, on progress in the creation of digital map data, this paper has attempted to summarise key factors in the definition of this task, and hence in the applicability of emerging technology. Some of these issues result in a strong divergence between the requirements of major users of map data and the requirements of the map production agencies. This divergence has to be resolved if Auto Carto is to produce, in the words of that deceptively simple maxim, "the right product, at the right time, at the right price."