International Hydrographic Review, Monaco, LXXI(1), March 1994

AUTOMATED ANNOTATION IN THE GEBCO DIGITAL ATLAS

by Gary J. ROBINSON [1]

Abstract

The GEBCO series of bathymetric charts of the world's oceans has developed over a long time, with many changes in format, content and detail. A digital version, the GEBCO Digital Atlas, is currently under preparation using the present (fifth) edition as a starting point. One of the methods for distribution will be the CD-ROM optical disk. With the potential of this medium for handling considerably more information than can be portrayed on paper and the use of the electronic medium for mapping applications, a number of cartographic issues are raised, one of which is feature annotation. Research into techniques for carrying out this process automatically has focused almost exclusively on land-based applications. The application of these techniques to bathymetric mapping is discussed, with particular attention paid to the problems involved with viewing data at arbitrary scale factors, and a number of possible solutions are presented.

INTRODUCTION

The GEBCO (GEneral Bathymetric Charts of the Oceans) series of oceanic bathymetric charts was established in 1903 under the auspices of Prince Albert the First of Monaco, a keen sailor and a "gentleman" oceanographer, who saw a need for accurate up-to-date charts of the ocean floors. The first edition was published in 1904 by the science committee set up by the Prince. Further editions were produced as additional data became available, and the International Hydrographic Bureau (IHO) eventually took over responsibility for GEBCO after the Prince's death. With the increasing scientific interest in accurate comprehensive bathymetry, particularly during the 1950s and 60s, the Intergovernmental Oceanographic Commission linked up with the IHO as joint sponsors of the project. The style of the latest version, the fifth, reflects this interest and differs substantially from previous versions, with

[1] NERC Unit for Thematic Information Systems, Department of Geography, University of Reading.

changes in the chart boundaries and the addition of trackline control and survey boxes to indicate reliability of the bathymetric contours. The GEBCO charts are compiled by international teams of oceanographers and bathymetrists using larger scale charts and other information. GEBCO is very much an ongoing venture because of the continual stream of new bathymetric information. That much of this arrives in digital form may change the way in which future charts are created.

The GEBCO Digital Atlas (GDA) is the culmination of a collaborative international effort over several years. Stable transparent copies of the colour separates used in the printing of the fifth edition were digitised by a number of organisations using a variety of manual, semi-automatic and fully automatic data capture techniques. The resulting data were checked and assembled to form the GDA by the British Oceanographic Data Centre (BODC), Birkenhead, England. The GDA contains the complete set of bathymetric contours with associated depth values, coastlines and trackline information. BODC currently distribute the data on magnetic tape, but the intention is to use CD-ROMs, which are gaining popularity as a cheap and efficient distribution medium. The importance of the GDA is demonstrated by its recognition as the logical "stepping stone" for the generation of the sixth and subsequent editions of the GEBCO series (GEBCO, 1991).

At present the GDA does not contain the names of undersea features that appear on the GEBCO charts. Instead, these, together with many more not able to be shown at the GEBCO scale of 1:10,000,000, are held in a separate computer readable gazetteer at the IHO in Monaco.

How these two products can best be combined, the issues involved in using computer displays for optimum presentation, and the possibilities and implications of adding non-bathymetric data, are the main subjects of this paper.

THE ELECTRONIC MEDIUM

Digital products such as the GDA have significant advantages over their paper counterparts. One of the more important is that they can be updated more easily and quickly. This is especially useful during the compilation process, and in the final product where new sections of bathymetry can be "patched in". Data in digital form are also more flexible. For example, subsets of the bathymetry, selected by specific contour values or ranges or by geographical region, can be extracted for use in a range of applications. Because the GDA is effectively scale-free and virtually unlimited in capacity more information can be stored, for example the names of small undersea features that cannot be shown on the current GEBCO charts, and other information such as marine geology.

Besides improving the selection and storage of bathymetric information the digital format also opens up a host of possible operations. Map projections other than those used in the paper charts (Mercator and Stereographic) can be applied according to requirements, whilst the contour data can be transformed into other products, for example gridded bathymetric models, which can be exploited in a wide range of scientific applications.

The flexibility presented by the GDA also creates new problems however, especially with the handling of text, such as feature names and contour labels. With the present format of the GDA, namely one file per sheet, simply reproducing the paper charts on a computer screen would lead to undesirable effects such as text being cropped at the screen edge when arbitrary regions are viewed. Multiple instances of the same label could also occur where the original sheet boundaries fall within the screen. Although the data within the GDA is currently held a one file per sheet basis it will eventually be possible to view or extract data by arbitrary region, perhaps straddling one or more sheet boundaries. For the purposes of manipulating data such as contours this is unimportant. However, feature names and contour labels are arranged on the original paper charts to suit the sheet limits.

A more serious problem arises from the fact that the GEBCO charts are at a fixed scale (1:10,000,000) and the information shown has been selected and laid out accordingly. This has implications for viewing the digital equivalent. The most important is that many computer screens have a resolution significantly inferior to that of the paper medium. At small scales the majority of the text would be illegible. Conversely, at large scales names would spread across the screen, with perhaps one or two large letters of a large feature such as an ocean being dominant. Some sort of scale dependent and position sensitive approach to the presentation of feature names is therefore required. One solution is to allocate to each name (or class of names) a scale range within which it should appear, coupled with a list of possible positions it could appear to cope with the sheet boundary problem. This method would allow names to appear both legibly and at sensible positions. The implementation of this approach by purely manual means would be impracticable however. This paper therefore concentrates on methods for the effective annotation of the GDA using automated means. These could be employed directly or "on-the-fly" as the data is displayed by the end user, or indirectly as discussed above, which is probably more practical.

THE GEBCO CHART SERIES

Before discussing the possible application of automated text placement techniques to the GDA it will be useful to review the design strategy used in the production of the paper charts, and how this relates to the characteristics of the ocean floors.

The technique used in the GEBCO charts for representing the morphology of the ocean floor is the isoline or bathymetric contour. The GEBCO chart is therefore a topographic product. The contours are compiled using larger scale charts such as those compiled in regional projects. Contours are notionally index or intermediate, although many intermediate contours are labelled in ambiguous situations, whilst some small index contours are not labelled. Contours surrounding depressions are denoted by interior ticks, irrespective of whether they are labelled or not. Index contours, which correspond to depths of 200m, 500m, 1000m and multiples of 1000m are also delineated by different colour infills. Intermediate contours are used in relatively flat regions to aid interpretation and so, with the exception of the 500m contour, are incomplete. Odd-valued intermediate contours are often hand-interpolated from even-valued intermediate contours taken from the source material. Dashed contours are used to indicate areas where the morphology is very uncertain. The treatment of areas where very little data are available, for example a single ship track may provide the only information, depends upon the compiler: some draw closed contours, others draw open ended (dangling) contours, whilst the remainder attempt to infer the continuity of the contours from more distant tracklines. This is the main reason for the variation in style of contours on GEBCO charts. To give some indication of the reliability of the contours a representative sample of the total number of ship tracks used in the compilation process are plotted as faint grey lines.

The other major items that appear on GEBCO charts are names of recognised topographic features, or groups of features. These are drawn from a catalogue of names approved by the National Committee on Undersea Names, a sub-committee of the Committee on Geographical Names. The font, size and orientation of these labels are used to indicate the type of feature: generally speaking, the larger the feature, the larger the text. Linear features such as ridges and trenches are generally labelled with the text aligned along their dominant trend, whilst small features such as sea mounts and areal features such as ocean basins and abyssal plains are labelled with horizontal text, located in a convenient but unambiguous position.

A characteristic of bathymetric nomenclature is the presence of a hierarchy in name classes. This reflects the natural hierachy that arises from the inter-relationships amongst features that lie on the ocean floor. For example, an ocean basin may contain sub-basins and seas, and each of these can contain continental shelves and abyssal plains, and so on. For large features such as oceans, this enables names of smaller features to be placed between adjoining letters of the name (Fig. 1). Names of small land features such as islands and island groups, which necessarily have to appear in the sea, are also placed using these conventions. Black ink is used to differentiate them from the names of undersea features, which appear in dark blue.

The positioning of depth labels along bathymetric contours also follows standard cartographic practices reasonably well. These are located in breaks in the contour lines, generally within the straighter sections of the contours where possible, and are oriented so that they can be read with the map held upright. Intermediate contours are labelled more often than hypsometric (land) contours, especially in areas of low slope where the index contours are widely spaced or where variations in topography would lead to ambiguous interpretations. The distribution of labels goes somewhat against the advice of IMHOF (IMHOF, 1982, p139) who points out that labels arranged in columnar stacks across consecutive contours can be visually detracting and do not perform their function effectively. However, it is interesting that many national land mapping organisations also ignore IMHOF's guidelines.

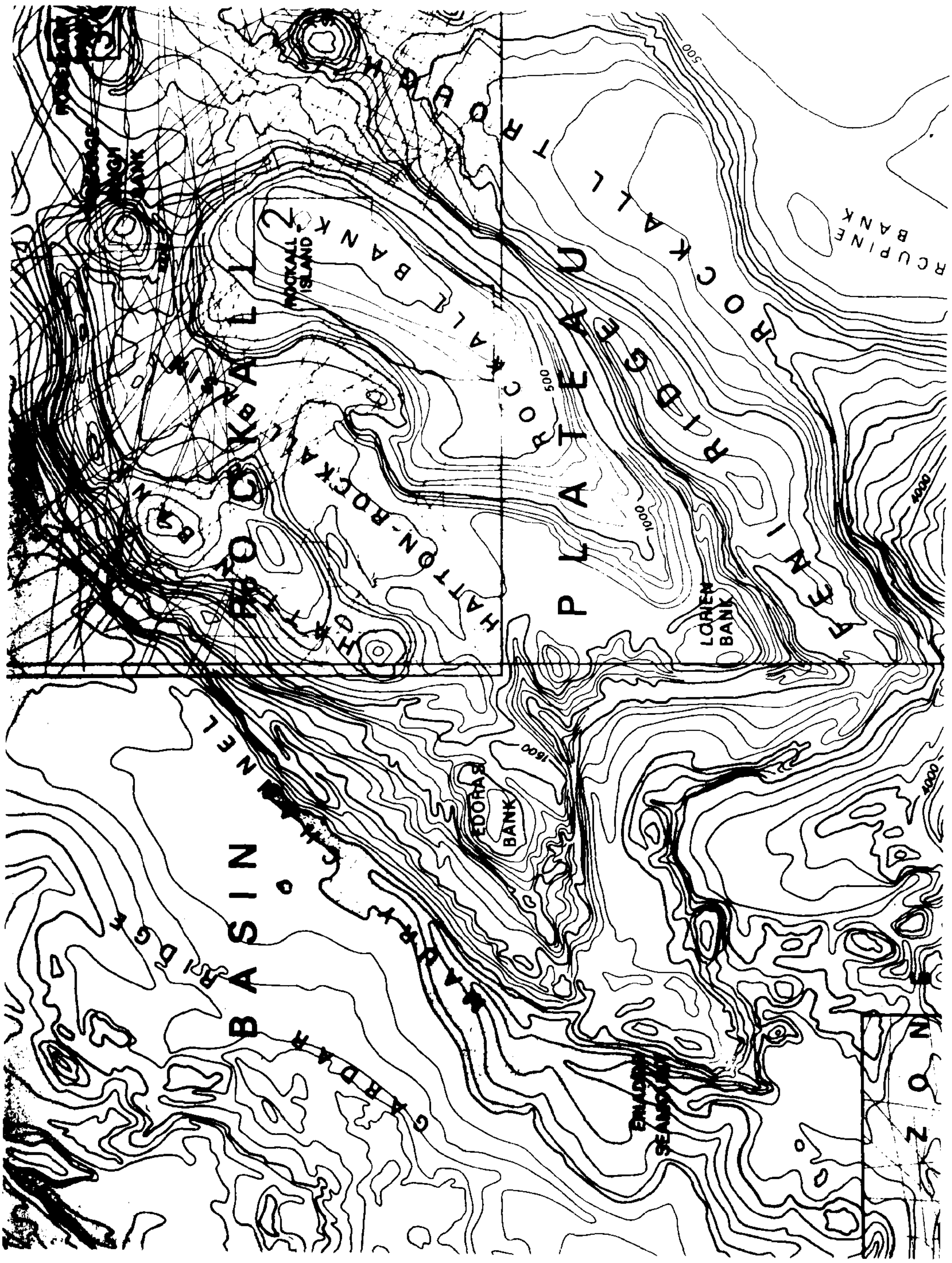

FIG. 1.- Section of a GEBCO 5th edition chart, showing text positioning.

THE GEBCO DIGITAL ATLAS

The nature of the digital data within the GDA differs from the paper charts in several respects. The most obvious is the absence of land contours and names of land features. This is because these are available in other data bases, e.g. Mundocart, or the Digital Chart of the World. Bathymetric contour depth values are not held as text items but as attributes of the digital contours. This allows considerable flexibility in the treatment of contours to suit the particular requirements of the user. For example, contours can be colour coded using a wide variety of schemes to indicate depth. Alternatively, subsets of contours, such as index contours, can be selected for display on the computer screen. These two approaches have been adopted in the GDA demonstration package produced by the IHO Data Center for Digital Bathymetry, based at the National Geophysical Data Center in Boulder, Colorado.

More sophisticated applications however will require the placement of contours labels on the screen, along with the names of undersea and other features. How this can be best carried out is the subject of the next section.

AUTOMATED ANNOTATION

Most of the literature on automated text placement or annotation relates to land based geographical applications, although together with the automated labelling of contours, the problem is more general. There are sufficient similarities between bathymetric charts and topographic maps to justify an examination of the approaches used in text placement in land-based maps and to learn from their successes and failures. Most of the approaches adopted are based on manual cartographic practices.

Most of the early work on automated text placement involved the labelling of point features such as towns and cities with horizontally aligned text strings of varying fonts and sizes (e.g., MOWER, 1986, CROMLEY, 1985). The problem is one of positioning the labels in order to maintain sufficient clarity (i.e. not overlapping other features or each other) but at the same time ensuring that they can be unambiguously related to their respective features by the map user. The essence of the problem therefore is to find an optimum set of positions for the labels according to some measure of best "fit". Most of the strategies proposed employ an iterative solution that attempts to minimise a function that is dependent on the various weightings given to possible positions. The success of these approaches hinges on whether a single solution exists for the set of positions. In general more than one exists, so the technique should recognise this. For point features a common technique is as follows:

Firstly, the points are ordered according to the number of degrees of freedom in placing their labels so that those that have the least amount of freedom are processed first. For example, points near the edge of the map are more difficult to label than those in the centre. Likewise, points in congested areas are more difficult to label than isolated points. The number of clashes or ambiguities between points and labels and between labels and labels is then calculated for each point in turn and the label position with the least conflict selected. This process is repeated until no clashes occur, or at least is reduced to an acceptable minimum. Similar techniques can be applied to labels of different sizes and orientations, although the computation of spatial overlap is more complex.

The labelling of a non-point feature is considerably more difficult because the label often has to be aligned with or span the feature. For linear features this means that the label has to follow the form of the feature: a river and its name is a good example. Long linear features often have the name repeated at intervals along their length to allow the map user to associate quickly the name with the feature and vice versa. How areal features are handled depends upon their size, shape and the scale at which they are shown: below certain thresholds small areal or linear features are treated as points, with consequent changes in the style of labelling (i.e. the label may be placed horizontally). Where an areal feature has no well-defined boundary, such as happens in bathymetric mapping, the constraints on the placement of the label are increased because the label itself may have to define the extent of the feature. How well the label spans the area depends upon the size and shape of the area and the text size of the label. Changing the size of the text and it's spacing, or repeating the label if the shape is very elongated or irregular, are common cartographic practices.

The process of automatically positioning a label (or labels) within an areal feature is relatively straightforward through the use of a 'medial axis skeleton' (FREEMAN and AHN, 1984). This is defined as the locus of a point that moves such that it maintains the maximum distance away from the inside edges of the area (Fig. 2). This produces a network (the skeleton), which can be pruned to leave a single branch over which the text can be positioned. This method can be generalised relatively easily to cope with situations where the area is clipped by the sheet edge, or in the case of the electronic medium, the screen edge. In this situation the edge forms part of the boundary of the area.

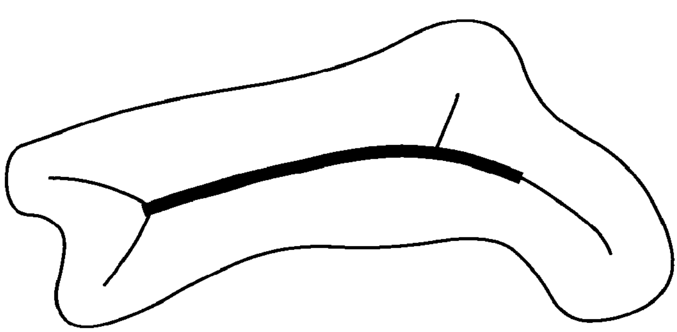

FIG. 2.- Medial Axis Skeletonisation of a polygon . The thick line is the result of "pruning" the skeleton.

By adopting a hierarchical strategy the annotation of mixed feature types can be split into a series of steps in which large areas are labelled first, followed by smaller areas and linear features, and finally small or point features. The influence of scale on this process can be illustrated using topographic maps of the United States. In this case the natural hierarchy of the various administrative units is exploited. At the country level only states and major cities will be labelled: the country name will be irrelevant and the counties will not be shown. Similarly at the state level the state name need not be shown as a label, although it may appear in the legend, and major cities and towns and county boundaries will be shown.

A common assumption in most studies of automated name placement is the existence of a set of preferred locations for the text relative to the point. This "cartographic convention" is based on very little published evidence, the main source being YOELI's model (YOELI, 1972) illustrated in Figure 3, in which the numbers refer to the order of preference in positioning. Recently, the rules for name placement have been reconsidered in the light of detailed studies of how this task is approached manually (WU and BUTTENFIELD, 1991). This study shows that the rather simplistic approaches adopted in existing attempts at automated text placement do not reproduce the manual process very well. Another interesting conclusion is that the manual process allows considerable flexibility in the placement of text on maps, especially those that contain areal features. For example, a town lying close to a lake may have its name overlapping the lake. The competition for space on the map is therefore not as great as may be first thought.

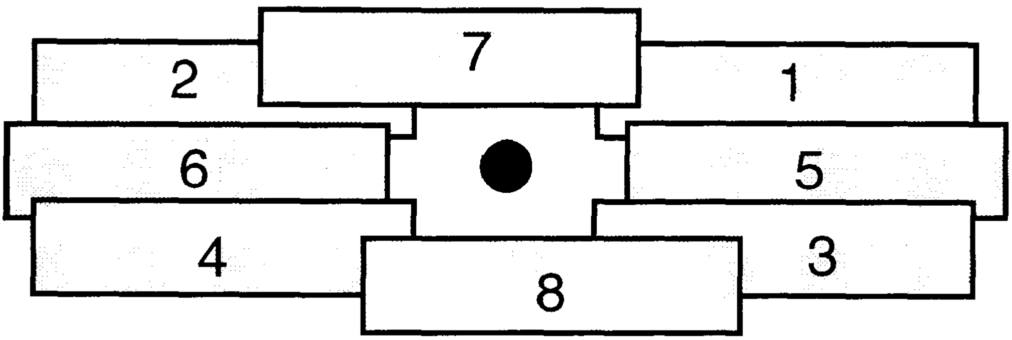

FIG. 3.- *Preferred positions of labels for a point feature, according to* YOELI *(1972).*

AUTOMATED CONTOUR LABELLING

Comparatively little research has been undertaken into the automated placement of contour labels in the context of topographic mapping. However, many commercial packages used in surface modelling, such as Uniras, provide options to label contours automatically. Some of the more cartographically-oriented systems, such as those marketed by Laser-Scan, Intergraph or SysScan, provide greater control over contour labelling. These allow the font and size of the labels and their spacing along contours to be selected. However, the relative spacing between labels tends to be constant and therefore does not allow for local changes in character along the contours or to the closeness of nearby contours.

The simplest method for labelling contours, which appears to be used in many commercial systems, is to compute (whilst plotting) the accumulated length of a contour line from its first appearance of the edge of the screen or plot. When this reaches one half of the selected interval plotting is suspended and the contour is further traversed for a distance equal to the length of the label. The positioning of the label within this gap can be carried out in several ways. The simplest method is to align the label along the vector from the start to the end of the gap (Fig. 4, left). A more sophisticated method is to fit a smooth curve, such as a spline, to the section of the contour defining the gap and "bending" the label to fit (figure 4, right). This latter technique is better for topographic maps in which a wide variation in curvature along contours is found. The process is then repeated at the selected label spacing until the contour exits from the screen. Because the last label may not fully lie on the screen, or occur within a short distance of the edge, some systems test for this eventuality and do not draw the label.

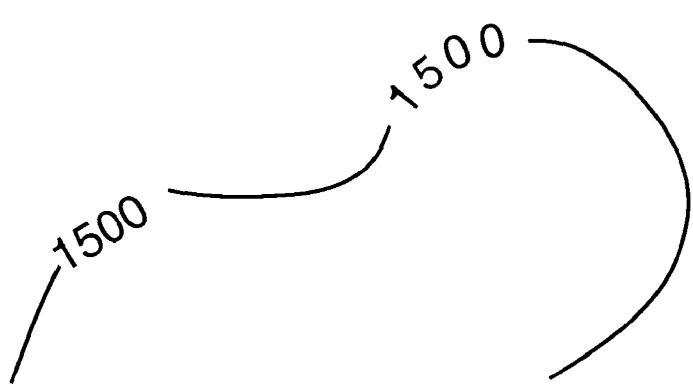

FIG. 4.- Applying depth labels to a contour. Left label uses a straight text string, right label uses a curved string.

Label orientation also depends on the system used. In general purpose systems labels are usually, but not always, oriented so that the parameter being contoured increases in value from the base to the top of the label, i.e. labels are placed "uphill". In cartographic systems however, labels are normally positioned so that they can be read with the map or chart held in an upright position.

APPLICATION OF AUTOMATED NAME PLACEMENT TO THE GDA

The similarities between land-based topographic maps and the GEBCO bathymetric maps suggests that the techniques developed for the former can be applied to the latter. In many ways the GEBCO charts are easier to deal with because of the fewer types of feature involved. The main difficulties foreseen in applying automated text placement techniques to the GDA however are more indirect in nature. The main one is that of successfully translating what is essentially a paper based atlas to an electronic equivalent. A whole host of visual design issues have to be addressed. For example, is a light on dark format best because this is what many people expect from a computer display, or should the display match as closely as possible the paper version? How should the wider choice of colour schemes provided by the electronic medium be exploited? Should colour be used to differentiate between contours? Other issues such as what level of detail to show or omit at particular scales, or even what additional information could be incorporated that is not presently shown on the paper charts, also have an impact on the implementation on an automated text placement scheme.

When only contours and names are required, as happens with the digital equivalent of the GEBCO charts, then the situation is somewhat simpler. Many of the existing approaches for mixed text placement can be applied with little change. What alterations will be required involve the handling of contour labels since this topic has not been really addressed, even by the comprehensive work of FREEMAN and AHN (1984). The actual mechanics of implementing an automated text placement system for the GDA will be the subject of a future paper. However, the development of a strategy for automatically labelling the GDA is reasonably straightforward and can be set out as follows:

1. The issue of what level of detail to show at a particular scale or geographic area has to be resolved. Labelling is one factor in this decision, since excessive detail on the screen or plot could result in insufficient room for labels to be placed properly. For contours this is a fairly simple decision: only index contours should perhaps be shown at scales of, say, less than 1:1,000,000. Another "rule" could be added that allows only 1000m contours to be shown at even smaller scales.

2. The handling of feature labels can be approached similarly. Each label could be assigned a scale break, below which it is not shown. Where individually named features form a named group the group name would then be shown instead. This approach would therefore exploit (and therefore rely on) the presence of a natural hierarchy in bathymetric features, and in particular, their names. Similar rules, but applied in the opposite sense, would cope with the labels of large features, which would not be shown at large scales. This approach results in a fairly simple generalisation scheme for feature selection and omission. (The question of whether the contour data should be filtered to provide "true" cartographic generalisation is not addressed in this paper).

3. The order in which labels are added depends upon how constrained they are by either the features they relate to, by neighbouring features, or by other labels. The general procedure would be to assign labels to the most constrained feature types, for example areal features, (e.g., ocean basins, seas, abyssal plains), then linear features (e.g., ridges, trenches), then small features (e.g., seamounts, knolls), and finally contours.

4. Conventional cartographic guidelines allow a certain degree of overlap between labels and some feature classes. In the case of the GDA (and as employed in the paper charts) this means that contours can be overlapped in places by text, since the hidden portion can generally be inferred by the shape of the contour (this works best for straight or moderately curved contour sections). Overlap of other feature types by text from another type of feature is also possible, but only where the labels can be assigned unambiguously (by the user) to their respective features, for example a seamount lying within an abyssal plain.

CONCLUSIONS

The application of automated text placement techniques to the GDA is technically feasible. However, a number of issues have to be resolved before this can be carried out. The main one is how best to transfer what is essentially a paper chart-based view of bathymetry to the electronic medium. The questions of what level of detail is required, especially with the potential for adding other information, also has to be addressed. Only when these problems are satisfactorily resolved can the existing techniques for text placement in land-based mapping be successfully applied to the bathymetric environment.

ACKNOWLEDGEMENTS

This work was undertaken under NERC contract F60/G6/12, and forms part of the NERC contribution to the GEBCO project. The assistance of other scientists involved in GEBCO, who made many helpful suggestions, is greatly appreciated.

REFERENCES

AHN, J. and FREEMAN, H., 1984, A Program for Automatic Name Placement, Proceedings, AUTO-CARTO 6, Ottowa, pp. 444-453.

CROMLEY, R. G., 1985, An LP Relaxation Procedure for Annotating Point Features using Interactive Graphics, Proceedings, AUTO-CARTO 7, International Symposium on Computer Assisted Cartography: Digital Representations of Knowledge, Washington D. C., pp. 127-132.

FREEMAN, H. and AHN, J., 1984, AUTONAP - An Expert System for Automatic Map Name Placement, Proceedings of the International Symposium on Spatial Data Handling, Zurich, Switzerland, Vol. 2, No. pp. 544-571.

GEBCO, 1994, GEBCO Digital Atlas - General Bathymetric Chart of the Oceans, International Hydrographic Office/International Oceanographic Commission/British Oceanographic Data Centre, Birkenhead, UK.

IMHOF, E., 1982, Cartographic Relief Presentation, (Translation of Kartographische Gelendedarstellung, edited by Steward, H. J.), Walter de Gruyter: Berlin.

MOWER, J. E., 1986, Name Placement of Point Features through Constraint Propagation, Proceedings of the Second International Symposium on Spatial Data Handling, Seattle, Washington, pp. 65-73.

WU, C. V. and BUTTENFIELD, B. P., 1991, Reconsidering Rules for Point-Feature Name Placement, Cartographica, Vol 28, No. 1, pp. 10-27.

YOELI, P., 1972. "The Logic of Automated Name Placement", The Cartographic Journal, Vol. 9, No. 2, pp. 99-108.

ZORASTER, S., 1991, Expert Systems and the Map Label Placement Problem, Cartographica, Vol. 28, No. 1, pp. 1-9.