9th Biennial International Conference
& Exposition on Petroleum Geophysics

"HYDERABAD 2012"

P-119

Decoding of CGM files for creating GIS based Corporate Database

Achintya Pal*, *ONGC*

Summary

Creation of a GIS based corporate database of Geological and Geophysical maps is a necessity of any hydrocarbon exploration company. However, many GIS databases do not directly accept the G&G maps which are normally generated and preserved as CGM (Computer Graphics Metafile) files. A procedure has been devised to covert these CGM plot files into ASCII files for portability of the G&G maps into the GIS Database. Even after the conversion to ASCII files, another important step need to be implemented before the data can be input to the GIS since the coordinates decoded from the CGM files are "device" coordinates. Accordingly, they have been converted to "user" coordinates like easting and northing (equivalently longitude and latitude, given the projection parameters) because it is only these coordinates that the GIS accepts, consistent with a chosen map projection system.

Introduction

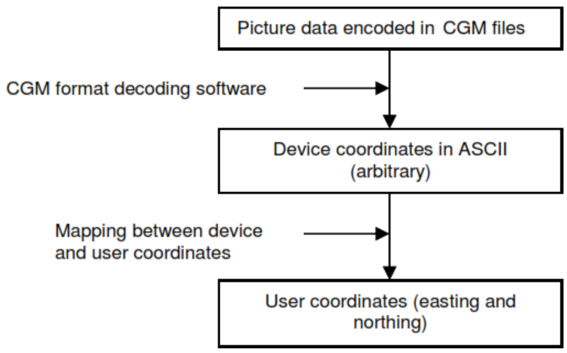
For the purpose of creating a organization-wide corporate database of all G&G maps, a user-friendly methodology has been developed to decode CGM plot files and convert them into ASCII files so that all the features contained in them can be imported into the GIS (Geographical Information System) database. CGM files are "picture" files which do not have an inherent "layered" structure that is required for creating a database. However, different map features like contours, faults, seismic lines etc. can be separated into layer-like structure using CGM file attributes such as line type, line width and line colour. The present work has achieved in deriving the original user-defined positional coordinates in an interpretation map in "layers" from the device coordinates hidden in the CGM files so that they can be stored in a GIS type database or can be utilized independently by any workstation user.

Methodology

The work can be divided into two essential parts:

(1) Decoding the CGM files to extract the arbitrary device coordinates of the picture file.

(2) Deriving the user coordinates (actual easting and northing on the map) by establishing correspondence with the device coordinates.



Step 1: Conversion of CGM file to extract coordinates in ASCII

Extraction of meaningful data from an encoded file requires specifications as to different codes and the location of bits and bytes of data. While pictures contained in CGM files can be drawn readily by any customized software on PCs and workstations, independent decoding of graphics metafile data was essential for the ultimate objective of this work, namely, finding the actual location of map features for wider use of the extracted data. The CGM format specifications are briefly described below.

A binary encoded CGM file consists of a sequence of "elements". The elements have a "class" number, "identifier" number (or element code) and a length (in bytes) indicator. The following elements are the minimal subset contained in a CGM file:

Regional Computer Centre, Oil & Natural Gas Corporation Ltd. Kolkata
babulan@gmail.com

Element name	Class	Identifier (code)	Remarks
Begin Metafile	0	1	*
Begin Picture	0	3	
Begin Picture Body	0	4	
Metafile Version	1	1	*
Metafile Element List	1	11	*
VDC (Virtual Device Coordinates) Extent	2	6	
Polyline	4	1	
Polygon	4	7	
Text	4	4	
Line Type	5	2	
Line Width	5	3	
Line Colour	5	4	
End Picture	0	5	
End Metafile	0	2	*

* These elements must be present in every CGM file.

A CGM file starts with a "Begin Metafile" element followed by a part called "Metafile descriptor". After the metafile descriptor follow zero, one or several pictures and finally the "End Metafile" element.

Begin Metafile	Metafile descriptor	Pictures …	End Metafile

The Metafile descriptor must contain at least two elements "Metafile Version" and "Metafile Element List". The parameter of Metafile Element List is a list of class and identifier codes of the non-mandatory elements that might appear in the file. After the Metafile descriptor elements, a sequence of pictures follows. Each picture has the following structure:

Begin Picture	Picture Descriptor	Begin Pic. Body	Picture element …	End Picture

The required graphical picture elements that may appear between "Begin Picture Body" and "End Picture" are Polyline, Text, Polygon etc. Polyline represents a sequence of connected lines and Polygon a set of vertices of a Polygon. The parameters of both are 2*p 16-bit signed integer values (length field: 4*p bytes) which are the VDCs of p points stored as pairs of X and Y coordinates.

Two forms of element structure are possible: a short-form element of parameter data lengths between 0 and 30 bytes and a long-form element for arbitrary parameter lengths. A short-form element starts with a 16-bit header of the form

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Class				Identifier (Code)							Length				

and is followed by the number of data bytes indicated in the 5-bit long length parameter. Long–form elements, with data lengths greater than 30, start with a 32-bit header of the form.

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Class				Identifier (Code)							1	1	1	1	1
Length															

followed by between 0 and 32767 bytes as indicated in the 16-bit length parameter. Bits 12-15 contain information on element class and 5-11 on element code.

Step 2: Mapping between device and user coordinate planes

After accomplishing the task of decoding device coordinates from the CGM file, a fairly accurate mapping of user coordinates (easting and northing) on to these device coordinates is very important for the purpose of deriving the actual geographical locations. This is done by establishing the correspondence of the lower left and the upper right corners on the device coordinate plane and the user coordinate plane. While the device coordinates are found out from the extracted data, the user coordinates are carefully read out from the G&G map. It is to be ensured that the ratios of height to width in both coordinates systems are same up to some desired accuracy for isotropic mapping between them, that is (refer to Fig. 1),

$$(yu2 - yu1)/(xu2 - xu1) \approx (yd2 - yd1)/(xd2 - xd1)$$

where (xd1, yd1) and (xd2, yd2) are the device coordinates of the lower left and upper right corners of a map and (xu1, yu1) and (xu2, yu2) are the corresponding user coordinates.

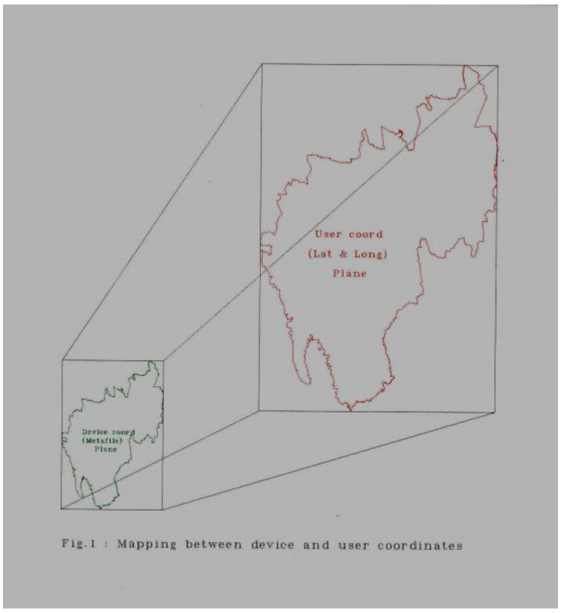


Fig. 1: Mapping between device and user coordinates

Results and Examples

Typical results for conversion of a CGM file to different features of a map ("layers" of information) are shown in Fig. 2 to 6. Fig. 2 is the plot of a typical G&G interpretation map stored in a CGM file containing contours, faults, seismic lines etc. After converting the contents of the CGM file to an ASCII file containing the actual positional coordinates of a G&G interpretation map along with attributes such as line type, line width and line colour, it is imported into the GIS. Fig. 3 to 6 depict the maps as they appear on the GIS display. Fig. 3 shows all the features together in a single frame. The objective is to separate them out in as many distinct features as possible so that they may be stored in the GIS database. The "layer" of contours and faults filtered out by the line colour attribute of the CGM file is depicted in Fig. 4. In a similar way, "layer" of bold contours, faults and some seismic lines selected by line width attribute of the CGM file is shown in Fig. 5. Finally, accuracy of the mapping between the internal coordinates of the CGM file and the actual easting and northing of the interpretation map is clearly indicated in Fig. 6 by superposing well locations from independent database sources on the well positions on the map.

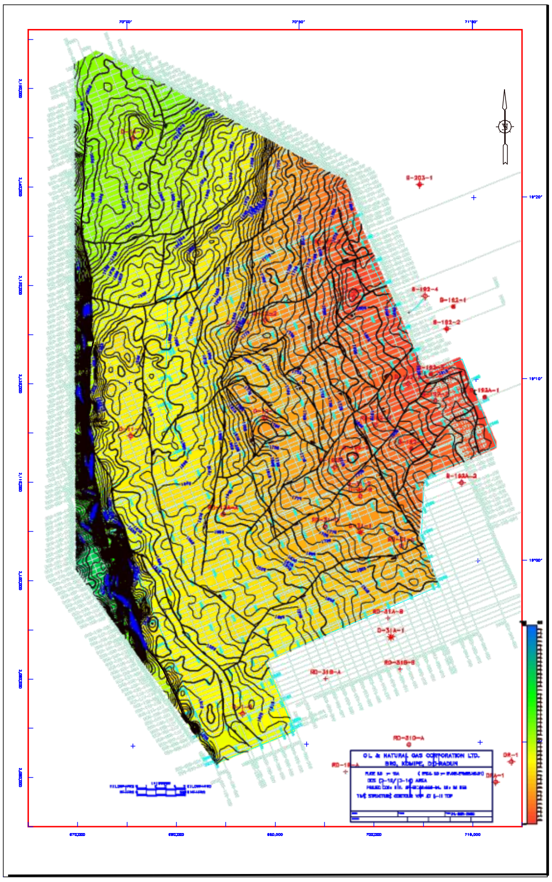


Fig. 2: Interpretation map from original CGM file

Fig 3: Interpretation map containing all features after import to GIS database

Fig. 4: "Layer" of contours and faults selected by "Line colour" attribute of the CGM file

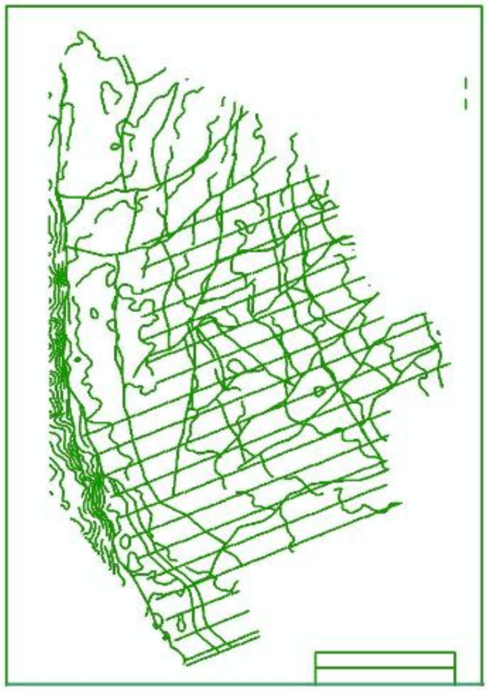


Fig. 5: "Layer" of bold contours, faults and some seismic lines selected by "Line width" attribute of the CGM file

Fig. 6: Accuracy of well positions superposed on well locations obtained from independent sources indicating proper mapping between device and user coordinates

Details of the Software

Representative flow-chart for the software is given in Section II. The software, written in FORTRAN, can be installed on UNIX based system as well as on PC. The name of the main CGM conversion module is "**decode_cgmfile**" and the subsequent module for transformation of "device" coordinates to the "user" coordinates is called "**dev_to_user**". Both executables are to be run from command line prompts with the following input and output specifications

Module 1: **decode_cgmfile**
Input: User specified CGM file name (Example: **Structure-Map.cgm**)
Output: User specified ASCII file name (Example: **Structure-Map.asc**)

Module 2: **dev_to_user**
Input:
(a) Lower Left and Upper Right corners (to be picked and read from output of the previous module e.g., **Structure-Map.asc**) of the device coordinate map in the **order xmin, ymin, xmax, ymax**
(b) Corresponding user coordinates of Lower Left and Upper Right corners of the map (to be read out from the hardcopy or softcopy of the map) in the order **xmin, ymin, xmax, ymax**
(c) Name of the decoded ASCII file (Example: **Structure-Map.asc**)
Output: User specified file name (Example: **Structure-Map.usr**)

From the study of a number of CGM files from different sources, it has been found that the map coordinates could be encoded in either of 2, 3 or 4 bytes. The module **decode_cgmfile** automatically detects this and proceeds to convert the contents of the CGM file so that the process is transparent to the user.

Conclusion and Remarks

A methodology has been devised to create a GIS-based layered database of G&G maps from CGM files in which final product of an interpretation project is stored, to facilitate corporate decision making. GIS does not accept the CGM files directly, unless the arbitrary device coordinates encoded in these files are decoded to ASCII (text data) and are subsequently converted to user coordinates. GIS requires a specific cartographic system for import of any positional data. All the map features are not only reproduced in their original easting and northing coordinates, they are given a database-like structure by using the attribute elements like line type, line width, line colour etc. (extracted from the CGM files) with which contours, faults, seismic lines as well as cultural features are represented in the original map.

It is, therefore, necessary while creating a CGM file that

(1) As many different line type, width and colour as possible should be used to represent different features on a map

(2) Map annotation box must contain all map projection related information. This helps in creating the cartographic parameters in the GIS database

(3) The easting and northing coordinates must be written along the boundary of a map along with the geographic coordinates and the fonts must be chosen in such a way that it is clearly readable on screen. This is crucial for a proper mapping of device to user coordinates (Step 2 of Section II)

Finally, it may be mentioned that apart from being put into a GIS database, the extracted coordinates and attributes can be imported back to an interpretation workstation whenever necessary. This will enable the users to re-create the map features with minimal effort.

Acknowledgement

The author wishes to express his gratitude to Mr. P.K. Bhowmick, ED, Head, KDMIPE, Mr. D.P. Sahasrabudhe, ED, Basin Manager, MBA Basin, Mr. B.K. Das, HGS and Mr. S. Panigrahi, GM, Head, RCC Kolkata of ONGC for promoting the work. The author is indebted to Mr. S. Chakraborty, DGM(Geoph) for his constant encouragement during the development of the software.

This work would not have been possible without the initiative of Mr. D.S. Mitra, DGM(Geol) of Remote Sensing and Geomatics Lab, KDMIPE and the tremendous help and cooperation given by him and his group members. The author is deeply indebted to them.