PICDMS: A GENERAL-PURPOSE GEOGRAPHIC MODELING SYSTEM

Margaret Chock, Ph. D.
Computer Consultant
1048 24th St.
Santa Monica, CA 90403

ABSTRACT

The Picture Database Management System (PICDMS) is a new system developed and implemented at the University of California, Los Angeles. It performs a wide range of data processing functions on sets of two-dimensional variables, as requested in a simple user language. The data base can be a set of pictures, images, or functions including, for example, digitized Landsat satellite photographs, an array of elevation measurements, and census tract boundaries for the same area, together with data derived for the region from tables or other sources. Possible operations include definition of new pictures such as vegetation maps based on elevation values and ground truth tables as well as the set of photographs; modification of existing images according to arbitrary criteria; calculation of the minimum distance from a geographic point to a specified census tract; deletion of an image; listing of sets of values derived from the image data base; computation of numeric information; or printing of maps showing correlations among the existing variables. New data can be added without affecting old user programs. The user gives the minimum information needed by the system for completely flexible operation; the system finds the rest of the information it needs in a data dictionary. The user specifies the type of operation (add, replace, or delete an image, compute a distance or some arbitrary numeric function, list values or print a picture). He defines the format of any new data, specifies the part of the region to be operated on in terms of values of the existing variables, and describes the effect on a set of variables of surrounding values of other variables. Simple operations can be defined in one- or two-line English-like sentences; arbitrarily complicated ones can also be defined in a straightforward manner. The system analyzes the user command in terms of the existing data base and the relative amount of computer memory available, and builds a simple computer program to perform the operation. A future Operations Dictionary will allow still easier definition of standard procedures.

INTRODUCTION

The Picture Database Management System PICDMS is a novel approach to two-dimensional data processing: an easy-to-use general-purpose computer system that allows a non-programmer to define simple or complicated operations over a surface such as the Earth, a medical image, or a silicon chip. Many

This research was part of a doctoral project chaired by Dr. A. Klinger and Dr. A. F. Cardenas of the University of California, Los Angeles, and partially supported by NSF grant MCS-7816754.

sets of data may belong to each surface, for example satellite photographs, maps, census tracts, etc. The user may perform standard and nonstandard operations on arbitrary subsets of this data, modifying or augmenting operations and data as he wishes.

This system was originally developed for applications in cartography and geographic modeling. Automated mapmaking is an example: registration of new aerial photographs to an old map, interpretation and abstraction of them to display new information, and overlay on the original map, all within a few hours. PICDMS allows experimentation with different output formats; different colors, symbols, and projections can be tried at little extra cost (as recommended by Morrison et al., 1975). Special-purpose maps can be produced from a general set of data in order to present a single theme clearly: data that is not immediately relevant is eliminated or downplayed. Such maps are often of too-limited interest to justify the expense of manual production. The same system can be used for very detailed, quantitative analysis of resources by comparing photographs to known ground truth information, as is done in Landsat multispectral grid cell classification. In addition, PICDMS can model geologic or geographic processes over a region using multiple types of data and providing snapshots at desired intervals. Manual computations for such two-dimensional models are necessarily very simple, both because of the large number of calculations and the time consumed in drawing maps. PICDMS permits more modeling steps, more experimentation, and more detailed and more complex models. Other possible uses of this software include brush fire modeling (Chock, 1982), and placement of health care facilities (Chock and Klinger, 1982).

DYNAMIC STACKED-IMAGE DATA STRUCTURE

The PICDMS dynamic stacked-image data structure is a set of images (which may be photographs, diagrams, or two-dimensional variables), registered to the same coordinate grid, with values for a single grid cell stored in a logical data record. The format of this record varies as images are added to or removed from the data base. The cellular rather than vector data structure provides superior representation of the subtle variations in photographs, and versatility in data manipulation. A variation on the "stacked image" concept proposed independently by Albert L. Zobrist (Zobrist, 1977) and by Robert M. Haralick (Haralick, 1977) is used: all images are registered to the same projection or camera location and scale, and a rectangular grid is then superimposed on them.

In order to allow interrelation of data from various sources, all data for a cell is stored in a single conventional data record (Figure 1). This format also allows a choice of data types for image values -- integers, floating-point numbers of various lengths, or character strings. The simple, sequential, uniform-format files can be used by virtually any computer system.

In other systems this structure is sometimes too rigid,

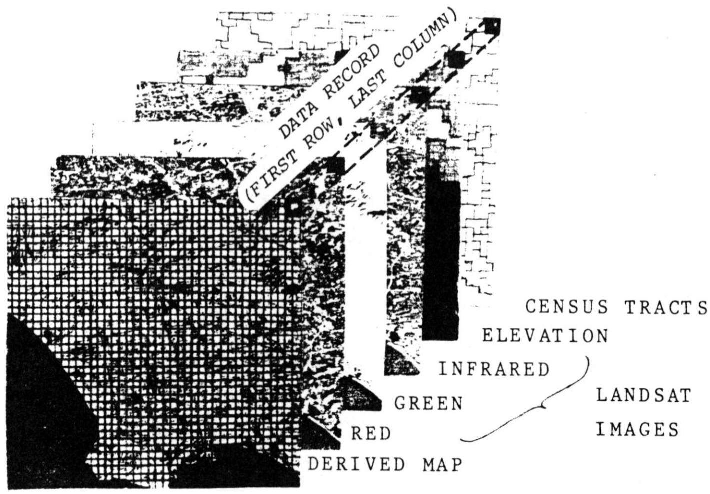

Figure 1. A Stacked Image Data Base. Each (row, column) record contains one data item per image: in this example, intensities for Landsat red, green, and infrared spectra, elevation, and pointers to census tracts.

requiring predefinition of the data: file management is simplified at the cost of storage wasted in unused blank fields; new attributes are required to conform in name, type, and length to the predefined fields; and system expansion is limited by the number of spare fields available for new data.

In PICDMS on the other hand the record structure is made variable: a new image is added as a new field in each data record, well-integrated with the existing data. A Data Dictionary keeps track of the current record format and job control language parameters needed by the system, by methods invisible to the users. There is no storage space wasted on blank images. The system can use a variety of image coding formats, and there are no limits on addition of new information to the data base.

A PICDMS data base consists of sets of these image stacks, each with its own area of coverage and scale. Each user sees his own private stack, to which new data he defines is automatically added, and which he can modify. He may also be able to read information from other stacks, but can not change them.

WINDOWED SCAN DATA MANIPULATION

The data manipulation facilities of PICDMS were built with two goals: a) to support any geographic data processing operation that could reasonably be performed on gridded data, and b) to make such operations as simple and standard as possible for the user without imposition of arbitrary restrictions on them. The first goal was set by collecting published data manipulation capabilities of a variety of automated geographic and cartographic systems (Table 1). This list is by no means exhaustive (some of

	KANDI DATS	IBIS	SYMAP	BASIS	IMAID	AGS	ODYS-SEY	POL-YVRT	NIMS	STAN DARD	WRIS	CGIS	GEO-QUEL	GADS
OPERATIONS ON IMAGES														
EDGE OPERATORS	G													
CLUSTERING	G													
COVARIANCE	G													
AVERAGE GRID CELL VALUES	G	G												
POINT CLASSIFICATION	G	G												
HISTOGRAMS OF GRID CELL VALUES	G	G	G											
THRESHOLDING	G	G			G									
TEMPLATE MATCHING					G									
OPERATIONS ON SCATTERED POINTS														
INTERPOLATION			G											
THEISSEN POLYGONS			G											
CONTOURS	G		G			P	P							
PROJECTION CHANGE	G	G				P	P	P						
OPERATIONS ON LINES														
NEAREST POINT ON A LINE					P									
CENTER OF A LINE					P									
GENERALIZATION							P	P						
INTERSECTION OF NAMED LINES					P		P							
POINT-TO LINE DISTANCES					P				P					
LINE LENGTHS					P				P	P	P			
OPERATIONS ON REGIONS														
POLYGON OVERLAY (INTERSECTION)	G	G	G	G	P		P		P		P	P	P	
CROSSTABULATION OF OVERLAYS		G							P					
FIND ADJACENT POLYGONS							P	P						
AGGREGATION								P			P	P		P
POINT IN POLYGON					P				P	P				
AREA					P					P	P	P		
CENTROID					P							P	P	
OPERATIONS ON REGION ATTRIBUTES														
RELATIONAL OPERATIONS					P								P	P

*P represents data in polygon format G represents data in gridded format

Table 1. Some data manipulation capabilities of cartographic and geographic systems (Chock, Klinger, and Cardenas, 1981)

these systems are known to have extensive libraries of programs), but it indicates some of the major capabilities useful in a cartographic system, and also the way the systems tend to specialize.

The PICDMS basic data access algorithm can be varied to support any of the data manipulation operations in Table 1, and many others besides. This algorithm runs a small window (Figure 2) over all designated regions covered by the image stack. Each cell in these regions has its turn as the "active" cell: relevant data from the cells surrounding it are read and manipulated, and the results modify the active cell, or add new information to it, or contribute to a set of calculations, or are listed or displayed in a picture.

The PICDMS Data Manipulation Language describes all necessary variations on the basic algorithm. It is a set of logic and arithmetic statements about values of any set of images as seen through the window, and specification of the use to which this information is to be put. Programs for the operations in Table 1 were written or sketched in this language. Some version of any of these operation types could be performed by at most two fairly simple commands.

The user specifies the type of operation (add, replace, or delete an image, compute a distance or some arbitrary numeric function, list values or print a picture). He

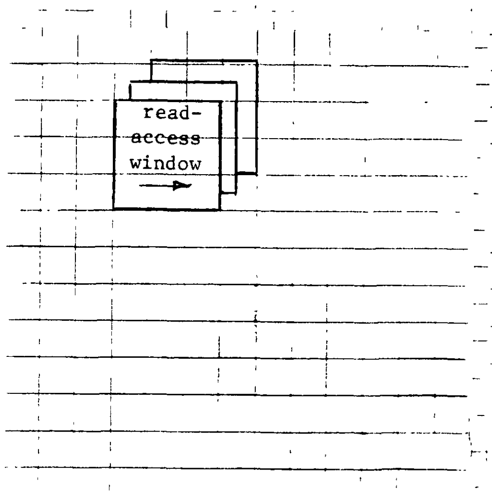

Figure 2: Windowed scan access to an image stack.

defines the format of any new data, specifies the part of the region to be operated on in terms of values in the existing images and describes the effect on a set of variables of surrounding values in a set of images. Simple operations can be defined in one- or two-line somewhat English-like sentences; arbitrarily complicated ones can also be defined in a straightforward manner.

The Data Manipulation Language will eventually support a family of simpler, more specialized languages for specific applications, with commands that can be defined and redefined for specific user groups.

EXAMPLES

The PICDMS Data Manipulation Language is perhaps best described by a set of examples.

Example 1

A vegetation map is derived from elevation measurements, ground truth tables, and aerial or satellite photographs:

```
ADD (VEGET IMAGE CHAR)
     VEGET = 0,
     IF BAND4 > 22 AND BAND4 <= 24 AND ELEV < 2000
          THEN VEGET =  'A',
     IF BAND4 > 24 AND BAND4 <= 30 AND ELEV < 2000
          THEN VEGET =  'B',
     IF BAND4 > 30 AND BAND5 <= 24 AND BAND5 > 18
          AND ELEV < 2000
               THEN VEGET =  'C',
     IF BAND4 > 30 AND BAND5 <= 30 AND BAND5 > 24
          AND ELEV < 2000
               THEN VEGET =  'D',
     FOR ELEV > 0;
```

Here a new vegetation map image called 'VEGET' is added to a stack which already contains spectral band 4 and band 5 photographs and an array of elevation measurements interpolated to the same pixel (resolution cell) locations. Vegetation type 'A' is known to reflect light at intensities between 22 and 24 at the band 4 wavelength, but only grows below 2000 feet of elevation. Other types are mapped according to similar sets of criteria. The symbol '0' is the default value if no vegetation type is defined for elevations above sea level; for lower elevations, the system leaves the map blank.

Example 2
The minimum distance from Los Angeles City Hall to Census Tract 900 is found:

```
DISTANCE FOR
    (LANDMARK = 'CITY HALL') AND (CITY = 'LOS ANGELES'),
    (CENSUS = 900);
```

The data base is scanned until a cell is found that fulfills one of the two conditions: there is a city hall in the 'landmark' image in a cell corresponding to the Los Angeles part of the image mapping cities; or the cell is in the '900' section of the Census Tract image. The scan continues until a cell is found satisfying the other condition. PICDMS calculates the distance between the two cells. The scan continues from the row and column where the second condition was first fulfilled. As more Tract 900 cells are found, their distances from the City Hall cell are calculated, and the minimum of these distances is printed.

Example 3
A set of values is listed:

```
LIST BAND6, BAND7 FOR VEGET = 'B';
```

The spectral band 6 and 7 light intensities are printed out for each cell with vegetation type B.

Example 4
The vegetation map image is discarded:

```
DELETE VEGET;
```

Example 5
The minimum elevation of Los Angeles County is found:

```
COMPUTE (V1 FLOAT), IF BEGINNING THEN V1 = 10000,
    V1 = MIN(V1, ELEVATION), FOR COUNTY = 'LOS ANGELES';
```

A floating-point number is initialized at a value known to be greater than the minimum elevation. During the raster scan, V1 is compared to the elevation of each cell in Los Angeles County, and the minimum value found is kept. After the scan, the minimum elevation value found is printed.

Example 6
The vegetation map of Example 1 is displayed rather than added to the data base:

```
PRINT, SYMBOL = 0,
     IF BAND4 > 22 AND BAND4 <= 24 AND ELEV < 2000
          THEN SYMBOL =  'A',
     IF BAND4 > 24 AND BAND4 <= 30 AND ELEV < 2000
          THEN SYMBOL =  'B',
     IF BAND4 > 30 AND BAND5 <= 24 AND BAND5 > 18
          AND ELEV < 2000
               THEN SYMBOL =  'C',
     IF BAND4 > 30 AND BAND5 <= 30 AND BAND5 > 24
          AND ELEV < 2000
               THEN SYMBOL =  'D',
     FOR ELEV > 0;
```

<u>Example 7</u>
PICDMS is unusual in its ability to operate in two-dimensional space on multiple variables, as in this version of the standard Roberts cross-operator for edge detection. The first PICDMS command below measures the band 7 spectral values from a Landsat image and creates an image ROBERTS, showing 'edges' where the gray level changes sharply.

```
ADD (ROBERTS IMAGE FIXED(10))
     ROBERTS = ABS(BAND7(I,J) - BAND7(I+1,J+1))
          + ABS(BAND7(I,J+1) - BAND7(I+1,J)),
     IF ROBERTS > 128 THEN ROBERTS = 1,
     ELSE ROBERTS = 0,
     FOR BAND7 NOT = BLANK;
```

The image added to the data base is called ROBERTS; it has ten-digit integer values. Each cell (I,J) of ROBERTS is assigned a value based on the value of BAND7 at (I,J) and at three neighboring cells. The operation is performed on all cells of BAND7 for which data is defined: undefined cells such as imaginary ones outside the image stack or those not assigned data values are described by the system code BLANK. The system will assign last row and column of ROBERTS the value BLANK, since values can't be computed by the given formula. If the resulting image shows too little detail in low-lying areas, the user might augment the command:

```
ADD (IMAGE ROBERTS FIXED(10))
     ROBERTS = ABS(BAND7(I,J) - BAND7(I+1,J+1))
          + ABS(BAND7(I,J+1) - BAND7(I+1,J)),
     IF (ELEVATION > 2000) AND (ROBERTS > 128)
          THEN ROBERTS = 1,
     ELSE (IF ELEVATION<= 2000) AND (ROBERTS > 64)
          THEN ROBERTS = 1,
     ELSE ROBERTS = 0,
     FOR BAND7 NOT = BLANK;
```

PICDMS SYSTEM CONFIGURATION

PICDMS uses a translator-processor system configuration to convert the current data base format and the user's immediate goals to an efficient computer process. The Translator (Figure 3) reads the user's command to determine the operation the user wants to perform, and on which data. It then reads the Data Dictionary to locate the required images and set up the input and output formats. It chooses an access method based on the operation, the quantity of data

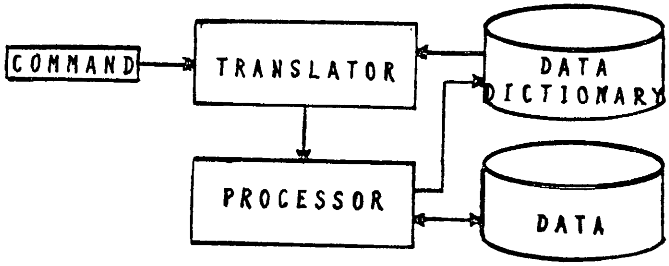

Figure 3. Translator-Processor Configuration

involved, and the amount of computer memory available.

There are currently three access methods; simple sequential raster scan of the image stack, whole-stack processing (if all stacks needed can fit in memory at once), and multirow (which allows efficient scan of each row of each stack for more limited memory, if raster scanning is inadequate for the operation). More direct access methods are also being developed, which again will be under system control. The Translator generates a Processor, a program which integrates all this information and does the actual work on the data base.

CONCLUSION

The current operational version of PICDMS is an end-user oriented, general-purpose processor of two-dimensional (particularly geographic and cartographic) data. It is being developed by addition of still simpler, more specialized languages, by increased capabilities (including interfaces to vector or line segment data, permanent storage of tables, and simultaneous use of data at different scales), and by adaptation to smaller computers and specialized processors (graphics devices, parallel processors). PICDMS is both a powerful, flexible tool for difficult data processing applications and a novel approach to data structure and access.

BIBLIOGRAPHY

Chock, M. I., 1982, A Data Base Management System for Image Processing, Doctoral Dissertation, Department of Computer Science, University of California, Los Angeles.

Chock, M., and Klinger, A., 1982, "A new approach to computer network cost modeling with medical applications," presented at World Congress: Medical Informatics and Developing Countries, Mexico City, 2/82, and to be published in Medical Informatics and Developing Countries, ed. by de Talens, A. Fernandez; Molino, E.; and Shires, D. B; North-Holland Publishing Co., Amsterdam.

Chock, M., Cardenas, A. F., and Klinger, A., 1981, "Manipulating data structures in pictorial information systems," Computer, Vol. 14, No. 11, pp. 43-50 (11/81).

Haralick, R. M., 1977, "Image access protocol for image processing software," IEEE Transactions on Software Engineering, SE-3 No. 2, pp. 190-192.

Morrison, J. L.; Ward, A. W.; and Devine, E. P.; 1975, "Implications of automation in cartography," The American Cartographer, V. 2, No. 1, pp. 82-88.

Zobrist, A. L., 1977, "Elements of an image-based information system," Workshop on Picture Data Description and Management, Chicago, 4/21-22/77, IEEE Computer Society 77CH1187-4C, pp. 55-60.