centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

SISTEMAS AVANZADOS DE GEO-INFORMATICA EN PLANEACION

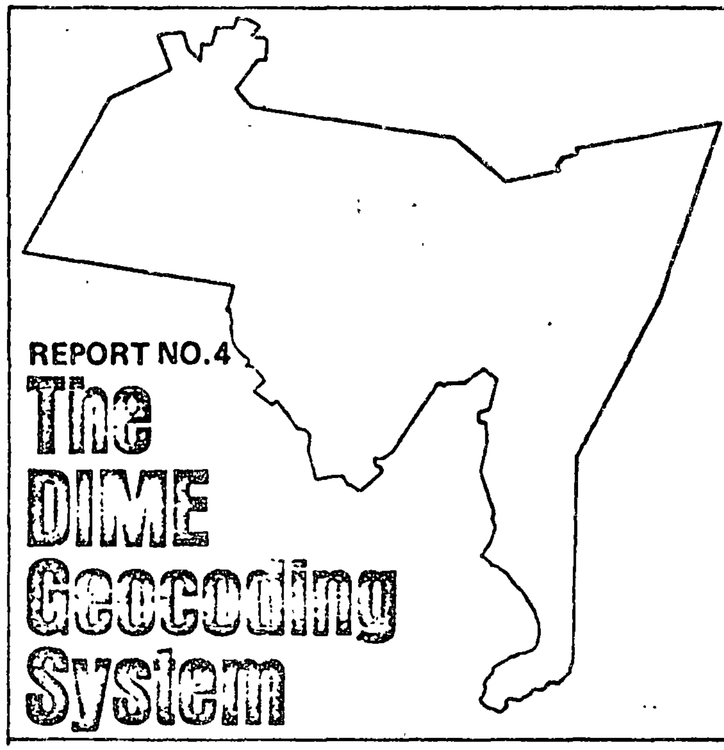
APLICACIONES URBANAS Y DE USO DEL SUELO

APENDICE BIBLIOGRAFICO

JULIO, 1978.

Palacio de Minería Calle de Tacuba 5, primer piso. México 1, D. F.

Census Use Study



REPORT NO. 4

The DIME Geocoding System

U.S. DEPARTMENT OF COMMERCE
Maurice H. Stans, Secretary
Rocco C. Siciliano, Under Secretary
Harold C. Passer,
Assistant Secretary for Economic Affairs

BUREAU OF THE CENSUS
George Hay Brown, Director

Issued July 1970

CONTENTS

Introduction

This chapter describes the Census Bureau's development of a computerized geographic system for use in conducting the 1970 census.

This report has been prepared to introduce and explain the DIME system to a three-part audience: (1) Individuals interested in technical developments in geographic coding, (2) organizations within standard metropolitan statistical areas (SMSA's) participating in or eligible to use the products of the Census Bureau's address coding guide/Dual Independent Map Encoding (ACG/DIME) program, and (3) organizations interested in creating their own DIME file.

For individuals interested in technical developments in geographic coding, this report describes the experiences of the Census Bureau in developing a computer file for relating addresses on census questionnaires to census geographic areas for tabulating the 1970 census. It describes early efforts to create such a file—the address coding guide (ACG) and the development of the Dual Independent Map Encoding (DIME) technique for creating geographic base files. It also describes the census geographic base file (GBF),[1] in which the ACG and DIME concepts are combined. The report also describes uses of a DIME file, creation of a DIME file, and the conceptual basis of the DIME system.

For those organizations within SMSA's which currently possess or are in the initial stages of creating a GBF, or those that will be able to obtain an already created GBF for their own use, this report describes the uses of the file in computer mapping, address and area coding, network and node analysis, adjacency analysis, and other uses. It also discusses the question of file maintenance and expansion.

For those organizations contemplating the creation of a DIME file, this report provides an overview of the system. In conjunction with a clerical procedures manual, a computer procedures manual, and various computer programs, this report provides the complete documentation needed to create a DIME geographic base file.[2]

1970 Census Procedures

In 1970 the Bureau of the Census conducted the Nineteenth Decennial Census of Population and Housing by a combination of two methods: a mail canvass in the large urban areas of the country and a house-to-house enumeration in the remainder of the country.

In 145 of the 233 SMSA's and in certain adjoining areas the mail canvass procedure was used to enumerate approximately 60 percent of the Nation's population. Instead of a visit to each household by an enumerator, census questionnaires were mailed to all residents. These questionnaires were addressed using a computerized mailing list derived from commercial

[1] The DIME file described in this report is essentially the same as the Census Bureau's GBF. Minor differences occur in formats and in lengths of particular fields, but the theoretical framework is identical.

[2] This report, the two procedures manuals, and the computer programs constitute the New Haven Census Use Study computer program package titled *DIME: A Geographic Base File System.*

sources and corrected and updated by the Post Office Department.

Before the census, addresses were checked by the Post Office Department. Addresses on cards were given to the postal carriers on each route with instructions to add, delete, or correct the addresses as required. Once corrected, the computerized list of addresses was used to generate mailing labels for the mail-out questionnaires. A second check was made at the time of delivery, with the necessary corrections in the listings and mailing pieces made manually.

For those portions of the mail canvass areas which were not covered by city delivery postal service, address lists were prepared by Census Bureau field personnel. These addresses were also corrected and updated by the Post Office Department at the time of delivery. This file of addresses was not computerized and the mailing pieces were addressed manually prior to the mail-out.

Immediately prior to census day—April 1, 1970—the questionnaires were delivered to all households. Householders were asked to complete the questionnaire and mail it to a local office. For those householders not mailing back the questionnaire or erroneously completing any of the items on the questionnaire, a followup was conducted by telephone or by personal visit by an enumerator.

The remainder of the country was enumerated by the traditional house-to-house canvassing procedure. In the urbanized portion of the 88 nonmail SMSA's and in those smaller cities which have contracted with the Census Bureau for block statistics, each questionnaire will be coded to its census block. In the remainder of the nonmail areas, the enumeration district (ED) will be the smallest geographic area to which a questionnaire is coded.

Geographic Coding for the Census

As in previous censuses, geographic codes for each household in the 1970 nonmail enumeration areas were determined and coded by the enumerator during the house-to-house enumeration process. In areas scheduled for block tabulations, the enumerator entered the census block number from a census map on each questionnaire during the enumeration process. For the balance of the nonmail areas, the enumerator identified the ED of each household and coded the ED number on the household's questionnaire. From a combination of codes for these "lowest common denominator" areas (ED's and/or blocks), codes for larger areas such as census tracts, townships, counties, and States were developed.

For the 145 mail census SMSA's, a method was needed to code individual addresses on the mailing list to specific geographic areas for tabulation purposes. A master coding file has been developed to code approximately 40 million mail addresses by computer to the appropriate geographic areas. The coding file developed is called the address coding guide (ACG).

The Bureau decided that the best way to create this massive file (or series of 145 files) was to—

1. Prepare a preliminary file in a standard format for each mail SMSA.

2. Send the individual SMSA files on standard coding forms to a local agency in the SMSA for review and correction, or review the file at the Census Bureau using available reference sources.

3. Edit the files by computer to obtain the most accurate files possible.

To create the preliminary files, the Bureau contracted with commercial firms for mailing lists and city directories. From these lists and directories and Bureau source files, computer records containing street names, address ranges for each block side, intersecting street names, and various geographic codes (such as State, county, congressional district, municipality, ZIP code, and census tract) were created.

To prepare the files for review and correction by either local cooperating agencies or by a centralized Census Bureau coding staff this information was computer-printed on FOSDIC worksheets.[3] Once prepared, the FOSDIC worksheets were sent to the

[3] FOSDIC (Film Optical Sensing Device for Input to Computers) is a coding method whereby prearranged small circles or spaces on a form or worksheet are coded using a "fill-in" coding method. FOSDIC worksheets, once coded, are microfilmed and the worksheet codes converted directly to codes on computer tape. In a loose sense the filled-in spaces on the worksheet serve the same purpose as punched holes in a punchcard.

cooperating agencies for review, correction, updating, and coding. The Department of Housing and Urban Development provided "701" funds to qualified local agencies to defray up to two-thirds of the local costs of this operation. Highway planning and research funds, administered by the U.S. Department of Transportation, were also available for this purpose.

For each block side within the coding area, the following codes were reviewed, and corrected or updated: Census tract, census block, beginning and ending address numbers,[4] area code (a municipality code), congressional district, voting wards (optional), and an optional or local code field. When review was completed, the worksheets were clerically edited by the local agency and a sample of the work was independently verified as a quality control measure. Units of work not passing quality control were reviewed and corrected. The completed worksheets were then returned to the Bureau for processing.

At the Bureau, the worksheets were microfilmed and the information on the worksheets was transferred to magnetic tape by the FOSDIC equipment. Using the information from the FOSDIC worksheets, the original file was updated. Computer edit routines were applied to the file and clerical corrections were inserted, as necessary. Coding errors were corrected by the Bureau staff when adequate reference documents were available. Other errors were referred to the local cooperating agency for correction. After edit and correction routines were completed for an SMSA, the file was made available to non-Bureau users at reproduction cost. Immediately prior to the census, the ACG files were updated from information supplied by the Post Office Department and by the commercial mailing list sources. The ACG files were then considered final and were used to code the census questionnaires geographically. They will also be available at the cost of reproduction to non-Bureau users. Figure 1 illustrates the basic steps necessary to create an ACG.

ACG files consist of block side records for all streets within the coded area. Nonstreet features, such as municipal boundaries, rivers, and railroad tracks are not included in the file. A block side is one side of a street between two intersections on that side of the street. A block side for a dead-end street is one side of the street from its beginning intersection to the dead end. Figure 2 illustrates some typical block sides.

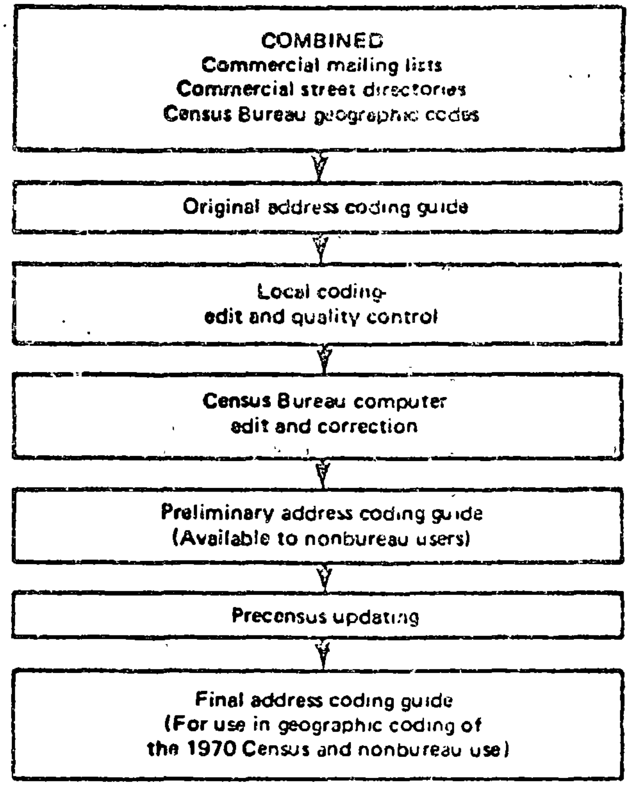


Figure 1. ACG creation process

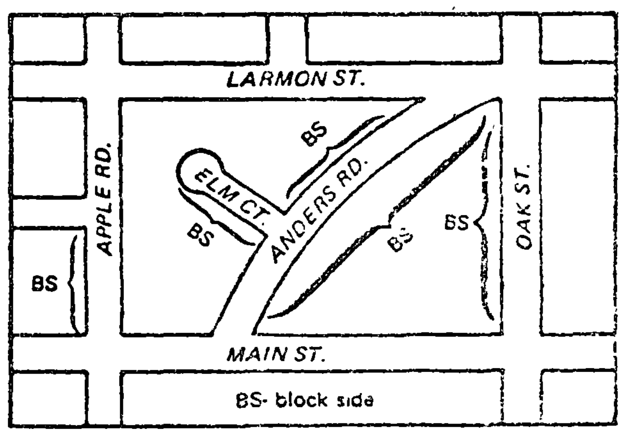


Figure 2. Typical block sides

Each ACG block side record contains the following codes: State, county, minor civil division (or census county division), place, ZIP code, 1970 census tract, street (includes street direction, name of street,

[4] For the purposes of this report, beginning and ending or lowest and highest address numbers refer to the lowest or highest *possible* address numbers, rather than actual address numbers, for each block side or street segment.

street type, and a street serial number), low and high address numbers, census block number, SMSA, district office (Bureau administrative code), area code (a code used in place of the minor civil division and place codes, above), optional code (local code or census serial number), ward (election district), annex (area annexed between the 1960 and 1970 censuses), congressional district, Post Office administrative codes, and serial number (unique number for each block side).

Development of DIME by the Census Use Study

The first ACG reviewed locally was the one for the New Haven, Conn. SMSA. This ACG was required to process the special census which was conducted in April 1967 to test the mail-out/mail-back procedures for the 1970 census. Local review and coding of the FOSDIC worksheets was completed in the fall of 1966.

The special census also provided a current data file for use in the research and development activities of the Census Use Study.[5] One of the primary research activities of the Use Study involved computer mapping.[6] A prerequisite for such research was the availability of a geographic base file with coordinates to permit the assigning of coordinates to data files for mapping. Since this type of file did not exist for New Haven, the Census Use Study attempted to create from the ACG a geographic file with coordinates.

Three alternative approaches were available for assigning coordinates to the ACG: (1) Assign coordinates for the approximate center point (centroid) of a block, (2) assign coordinates for the centroid of a block side, and (3) assign coordinates for both ends of each block side. The third alternative was selected because it provided the capability of plotting a street or network map that could be used for display or editing purposes. It also provided the ability to map by shading data values within the actual configuration of a block. The other two alternatives did not provide these capabilities. However, the selection of the third alternative made it possible to compute coordinates for the other two approaches, i.e., block centroid and block side centroid.

---

[5]See Census Use Study report, *General Description.*
[6]See Census Use Study report, *Computer Mapping.*

Portions of the New Haven ACG were tested using the third alternative. Coordinates for the beginning and ending of each ACG block side as shown on the Census Bureau's single-line metropolitan maps were measured and recorded using a semiautomatic device called a "Coordinate Locator" built by the Bureau. This process is called "digitizing." Coordinates digitized for each ACG block side were converted to state plane grid coordinates and added to the ACG file as part of each ACG block side record.

Substantial problems were encountered during the testing of the digitizing process. Since the ACG is a file of streets only, nonstreet features existing in the coded area such as railroad tracks, rivers, lake shores, and municipal boundaries were not included in the file and therefore could not be digitized. The absence of nonstreet features which formed block boundaries made it impossible to accurately plot block maps or calculate the area of blocks. This absence also limited the usefulness of a plotted map for display or analysis.

Another problem was that the digitizing process itself was inefficient: for the typical intersection (two streets crossing), eight readings had to be made on the coordinate locator, one for each of the block side ends forming the intersection. When such an intersection was plotted, the plot usually showed more than one of the plotted points at the intersection in slightly differing locations. This discrepancy was due to variability introduced manually in digitizing the point eight times and also to error introduced by mechanical slippage in the digitizer.

Still another problem was the inability to digitize curved streets accurately. In the ACG, each block side is coded from intersection to intersection without regard to curves that may occur along the block side. As a result, when a curved block side was digitized and plotted, it plotted out as a straight line. This occasionally led to plotted curved streets crossing other plotted streets incorrectly and the depicting of nonexistent intersections.

A substantial number of coding errors and omissions were detected through the New Haven ACG test as it was used by the Census Use Study. The discovery of these errors, not detected by the existing ACG clerical and computer edits, and the coordinate assignment problems discussed above, led to the conclusion that alternative approaches should be explored. Although the ACG was sufficiently accurate for the geographic coding of census questionnaires, the Census Use Study concluded that improving it for other purposes, such as computer mapping, would not be practical.

The problem of developing a new system was referred to the Technical Steering Group (TSG) of the Census Bureau Advisory Committee on Small Area Data, the body overseeing Census Use Study activities. A subcommittee was established to explore other methods of automated geographic coding or to conduct original research, if necessary, to develop a geographic base file which could be used efficiently for computer mapping and geographic analysis.

The subcommittee first considered extant methods of creating geographic base files. Then it developed a proposal for constructing a geographic base file using graph theory as the conceptual framework. Since a linear graph can be described as a series of straight line segments in a plane connected at vertices, it is readily apparent that a single line map is a form of linear graph. Figure 3 illustrates this point.

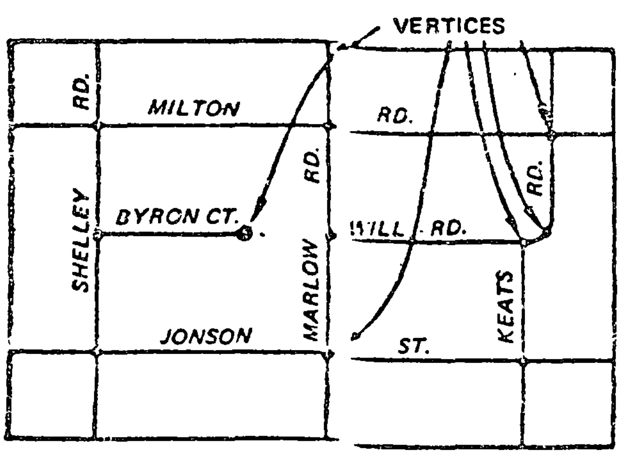


Figure 3. Graph elements of a single line map

Each street, river, railroad track, municipal boundary, or other map feature can be considered as one or more straight line segments. Curved lines can be divided into a series of straight line segments. When features intersect or where straight line segments change direction, vertices are formed. Using a concept derived from graph theory, a method was devised for creating a geographic base file where each vertex, intersection, or node,[7] as well as each line segment and intervening enclosed area, could be uniquely identified. Also each could be identified in terms of its place within the entire region encompassed by the file and in relation to adjacent nodes, line segments, and enclosed areas. The entire region encompassed by the geographic base file could be viewed as a series of interrelated nodes, lines, and enclosed areas. Each line segment is identified by its name as taken from the source map, e.g., Main Street or Green River, or by the node numbers at each end of the line segment. Nodes are numbered sequentially and uniquely. Enclosed areas are numbered using a systematic series of block numbers or other areal identifiers, or they can be described by the nodes or line segments which bound them.

The creation of such a file of line segments, nodes, and enclosed areas allows for the editing of the file by computer using an algorithm derived from graph theory. When coded, the three elements—segments, nodes, and enclosed areas—can be formed into separate incidence matrices, e.g., line segment-node, line segment-enclosed area, and enclosed area-node, by computer. The relationship established within these incidence matrices provided for the development of the computer editing system.

Such a file, as a representation of a map, can be digitized and plotted by computer to produce a complete replica of the source map. Often it is impractical or unnecessary to correct every edit reject or to reproduce faithfully all curved features. However, a perfect file can be produced at a reasonable additional cost.

The system that developed from this research activity was named Dual Independent Map Encoding (DIME) because the basic file is created by coding two independent incidence matrices from the source map.

From this theoretical base, the Census Use Study staff developed procedures and created a DIME file for New Haven. The same type of reference sources used to create the ACG—Census Bureau metropolitan

[7] The terms vertex, intersection, and node are considered to be essentially the same for the purposes of this paper, node is the term used because of its wide use in transportation planning and other geographic applications.

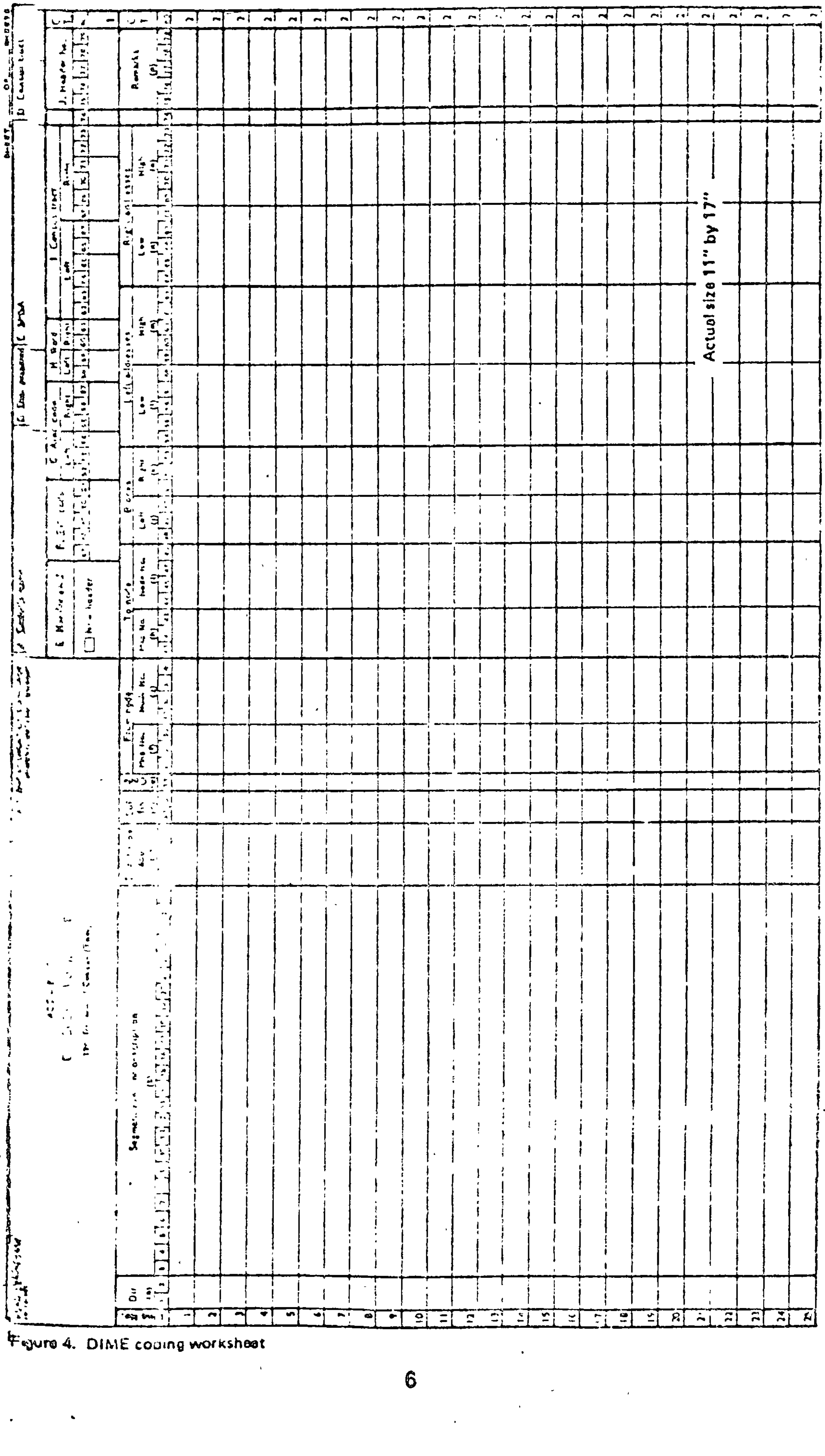


Figure 4. DIME coding worksheet

maps and local address maps or listings—was employed. Only one major change was made to the metropolitan maps prior to DIME coding—each node on the map was assigned a unique identification number. The coding form used was a standard punchcard transcription worksheet. A replica of the final coding form developed for the system is shown as figure 4.

While an ACG is constructed on a block-side basis, a DIME file is constructed on a street-segment basis. Each ACG record contains the appropriate codes for *one* side of a street between two intersections on that side of the street. Each DIME segment record contains the appropriate codes for *both* sides of a street between two nodes.

Essentially the same information is contained in both systems—street name, address number ranges, census block numbers, census tract numbers, and other geographic codes. The DIME system has two additional codes: node numbers and left/right orientation.[8] The DIME system also contains all meaningful nonstreet features such as rivers, municipal boundaries, shorelines, and railroad tracks shown on the metropolitan maps. These are not contained in the ACG system. Nodes are placed at sharp curves in streets or other features so that such curves can be adequately described by a series of straight line segments when plotted by computer.

Worksheets were coded and keypunched; the data were put on magnetic tape. The file was computer edited and edit rejects were corrected. After two or three edit cycles, the file was considered sufficiently accurate. The metropolitan maps used to code the DIME file were then digitized and the node numbers and coordinates were merged into the DIME file. The file was then ready for use in computer mapping.[9] The file was also used for address and area coding, as discussed in chapter II.

Further Experiments

To examine the applicability of DIME to ongoing programs of the Bureau, and to develop standard coding procedures and Bureau capabilities in the use of the system, further research was initiated.

A coding manual, forms, and procedures based on the New Haven experience were developed; and Roanoke, Va. was chosen as a test site for further DIME development activities. The Roanoke City Planning Commission agreed to supply coding clerks. The Use Study staff trained the clerks and supervised the coding operation.

After the local coding operation was completed, the coding worksheets were keypunched and the punchcards were converted to magnetic tape. The file was edited using DIME edit techniques. Rejects were corrected and inserted in the file. The coding maps were digitized with the Bureau's digitizing equipment. Nodes and coordinates from this process were inserted into the DIME file. Test maps were prepared on the Bureau's Calcomp plotter. Coordinate errors detected on the maps were corrected and the file updated. The file was then considered complete.

Further testing of coding procedures was conducted at Binghamton, N.Y., by the Broome County Planning Board. Areal coverage for this test was extended beyond the central city to the county boundary.

After training by Census Use Study staff, local personnel completed the coding operation with only nominal monitoring by Census Use Study staff. The Census Bureau's metropolitan maps were used for the urbanized portion of the county. County highway maps were used for the remainder. The manuals and worksheets used at Roanoke were revised and a coding manual was introduced for this test.

When coding and quality control operations were completed, the worksheets were keypunched and computer edited. The maps used for coding were digitized. When completed, the coordinate information was merged into the DIME file at the Census Bureau. Test maps were plotted and the file was corrected to complete the Broome County test.

---

[8] The left/right orientation code separates the codes for one side of a line segment from the other.

[9] The computer mapping experiments conducted are described fully in the Census Use Study report, *Computer Mapping*.

The final test leading to the development of the Census Use Study's computer program package, *DIME: A Geographic Base File System*, was conducted in the Greenville, S.C. SMSA by the Greenville County Planning Commission. Prior to the test, the coding and supervisors manuals and the coding and administrative forms were refined according to findings of the previous tests. However, this test was designed not only to evaluate the DIME computer program package but also to develop materials which could be used to create geographic base files for the nonmail census SMSA's.[10] As a result, materials which could be adapted to both uses were developed. Geographic coverage was limited to the city postal delivery portion of the urbanized area of the SMSA, as in the ACG program.

As in Binghamton, supervisory training was conducted by Census Use staff members. The supervisor, in turn, trained the coding staff. Biweekly monitoring visits were made by Census staff. Upon completion of coding and local quality control measures, the worksheets were returned to the Census Bureau. Keypunching, editing, and corrections were completed, and the file became the prototype for the nonmail census ACG/DIME geographic base file program.

Concurrent with the Greenville test, an experiment was conducted to examine the feasibility of merging an independently created DIME file without addresses for Madison, Wis. with the ACG for Madison. The DIME file was created and edited by the Use Study staff; the ACG file was coded by a local agency and processed at the Bureau as part of the standard ACG program. The metropolitan maps were used as the base for creating both files. Theoretically it should have been possible to merge these two files into an ACG/DIME file. However, a substantial number of records could not be merged. For example, of the 2,345 block records which could be merged, 1,104 records had at least one difference from one file to the other file. An analysis of the differences revealed that 19 percent were street name spelling differences, 3 percent were DIME errors, 47 percent were ACG errors, 9 percent were errors in both files, and 22 percent were ambiguities in the metropolitan maps. Reconciling the various differences took a substantial amount of time. In fact, more time was spent on this process than would be required to create a DIME file with addresses for Madison using the Greenville approach. As a result, the merging of separate ACG and DIME files by computer was discarded.

However, the desirability of inserting DIME features into the ACG's increased as it became apparent that the ACG's could contain a considerable number of errors that were not detectable by existing ACG edits. The benefits of the DIME edits and the potential uses of DIME files for computer mapping and other uses were also becoming widely known. Inquiries as to the availability of DIME files in the mail census SMSA's were received by the Bureau from local, State, and Federal agencies.

After the close of the ACG/DIME merging experiment, further research into adding DIME features to the ACG's by computer was undertaken by both the Census Use Study and the Bureau's Geography Division. The Geography Division experimented with inserting unique, computer-generated node numbers into a specially prepared ACG. The attempt was only partially successful and was abandoned.

At approximately the same time, the Census Use Study investigated the possibility of adding DIME features, i.e., node number and left/right parity status, to a specially prepared ACG block side file by clerical procedures. The following procedure was developed for adding these features:

1. Metropolitan maps would be node numbered as in DIME; i.e., all intersections and points along curved features would be numbered uniquely.

2. A specially formatted worksheet printout version of the ACG allowing for the clerical insertion of node numbers and left/right parity status for each block side in the file would be prepared.

3. Nodes and parity status would be coded on the worksheet. Additional records would be coded for nonstreet features (nonstreet features were not coded in the ACG) and for block sides which had to be split into two or more pieces because of curves or other reasons.

4. After coding was completed, the added codes (and unique identification numbers for each record) would be keypunched and merged into the original ACG file. The resulting "segment side" records would then be merged to form segment records.

[10] Nonmail census geographic base files will be discussed in greater detail in the next section of this chapter.

5. The resulting file would be edited using the DIME edits.

6. Edit rejects would be corrected.

7. The metropolitan maps used in coding would be digitized to create a file of node records—each node record would contain the node number and its coordinates.

8. The node record file would be merged with the DIME file and would thus be in final form for release to appropriate agencies.

A generalized flow chart of this operation is shown in figure 5.

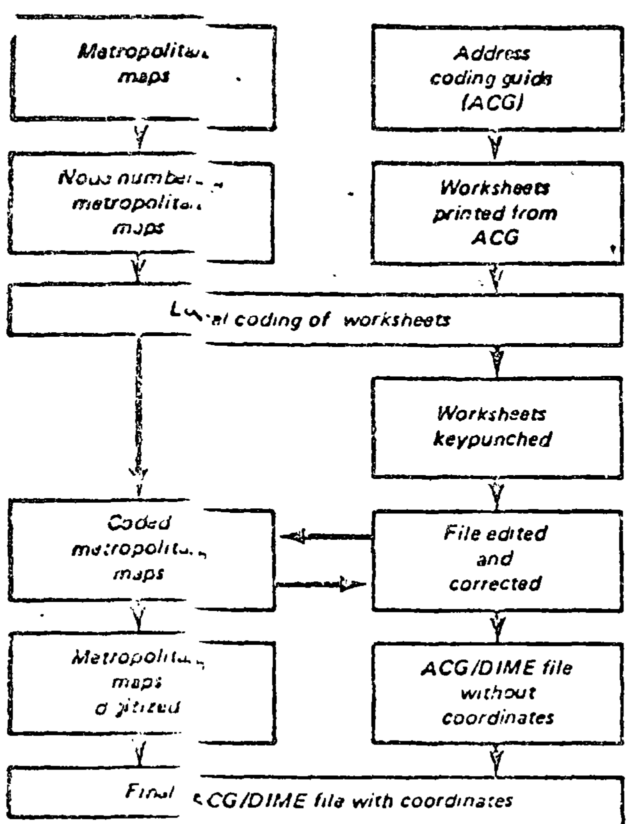


Figure 5. ACG/DIME creation process

This technique was tested using the ACG for a small number of census tracts in Madison, Wis. Testing showed the system was workable and sufficiently economical to warrant its use in a nationwide program. The Bureau accepted it for the program for inserting DIME features into the ACG's.

When the technique was accepted, a task force was created to conduct a large-scale test and prepare necessary clerical manuals and computer programs. This test was conducted successfully in Madison, Wis. under field conditions using draft manuals. The file was processed using the steps outlined above. The ACG/DIME program was then turned over to the Bureau's operating divisions for implementation.

The ACG/DIME Geographic Base File Program

As discussed earlier in this chapter, the 1970 census is being conducted using two different enumeration methods. In the major urban areas of the country (approximately 60 percent of the population), the census will be conducted using the mail-out/mail-back technique. The remainder of the country will be enumerated by traditional house-to-house canvassing methods. The Bureau has established one geographic base file program for the 145 mail census SMSA's and a separate program for the 88 nonmail census SMSA's. Different methods are being used to create ACG/DIME geographic base files in each program, although the completed files are essentially identical. In the mail census areas, the ACG's are being "improved" by adding DIME features. In the nonmail census areas, ACG/DIME files are being developed "from scratch." A listing of the mail and nonmail census SMSA's and the availability of ACG and ACG/DIME files is provided in appendix A.

Mail Census Geographic Base File Program

For the SMSA's included in the mail census, ACG's have been prepared for the central city (or cities) of the SMSA and the surrounding urbanized area, generally for the portion of the SMSA in which mail is delivered to a street address (such as 115 Main St.). The Bureau has required this areal coverage because each census questionnaire must be coded to its proper geographic area for tabulation purposes. Geographic coding of the questionnaires is accomplished by computer matching the address to which a questionnaire is mailed (115 Main St.) to its matching address range in the ACG (101-199 Main St.). Once this match is accomplished, the requisite geographic codes in the ACG are appended to the computer-generated record for the questionnaire.

The ACG/DIME geographic base file program in the mail census SMSA's (called the ACG Improvement Program) originated when it became obvious to the Bureau, and other Federal, State, and local agencies, that it would be desirable to add DIME features to the already existing ACG's. Federal agencies, including the Department of Housing and Urban Development and the Department of Transportation, became interested in supporting the program. They provided financial support when the Bureau decided to proceed with the program.

Time requirements for creating the machinery to conduct the 1970 census will not allow the use of the improved ACG's. The ACG/DIME files will, however, be used in future censuses and surveys, and will be periodically updated.[11] The files are being developed primarily for local and State agencies concerned with small area data analysis and planning. They will also be useful for a variety of planning and analytical studies in the private as well as the public sector. Uses are discussed in chapter II.

The local coding phase of the program is scheduled to conclude in late 1970. In the majority of SMSA's, coding should take from 3 to 6 weeks. However, the larger SMSA's may take from 3 to 4 months. Computer editing will begin when the coding worksheets are returned to the Bureau. The processed geographic base files (without coordinates) are scheduled to begin becoming available 6 to 9 months after the coding worksheets are returned to the Bureau. They should be available for use soon after the census summary computer tapes begin to become available. Although the census will be conducted using the presently existing ACG's, the geographic area codes—census tracts, census blocks, street names, address ranges, etc.—will be essentially the same in both the ACG and the improved ACG/DIME files. Therefore data from the 1970 census will be usable with either file.

Coordinates for the ACG/DIME geographic base files are scheduled to be digitized from the node-numbered coding maps beginning in calendar 1971. The coordinates are scheduled to be added to the geographic base files, and tested for accuracy.[12] The files will be available on a flow basis beginning in late 1971 and continuing through 1972.

Nonmail Census Geographic Base File Program

In nonmail census areas, place-of-work addresses reported on the sample census questionnaires will be coded to census tracts for tabulation purposes. To do this, an address coding guide that relates addresses to the census tract level is required. The Bureau intended to create only census tract address coding guides for this purpose. Since such guides would be of little use to local users or to the Bureau in future censuses because census tract boundaries often change between censuses, the Bureau decided to develop DIME-type geographic base files for the nonmail SMSA's when the Department of Housing and Urban Development and the Department of Transportation agreed to support a program to create such files.[13] Nearly all of the SMSA's eligible for the program have participated (see appendix A). Census tract coding guides will be prepared for those SMSA's not participating.

The process for creating an ACG/DIME geographic base file in this program is essentially the same as the method developed in New Haven and refined in subsequent tests. The local coding phase of the program was completed in early 1970. Computer processing and editing of the files is scheduled to be completed by mid-1970. Files without coordinates will become available for local use at that time. The coding maps are scheduled to be digitized during 1971 and files with coordinates will be released on a flow basis to local agencies starting in spring 1971 and continuing into 1972.

---

[11] The Southern California Regional Information Study (SCRIS), a joint research study sponsored by the Bureau and the Southern California Association of Governments, of which the Census Use Study forms the Bureau's contingent, has been charged with developing a maintenance and updating system for the ACG/DIME Geographic Base File Program. Development of the system is scheduled to be completed during 1971.

[12] Coordinates in the ACG/DIME files will be expressed as geographic coordinates (expressed in degrees of latitude and longitude carried to four decimal places), state plane coordinates and "map miles" from an arbitrary point.

[13] Funds have been made available for local agency coding work under the "701" program of the Department of Housing and Urban Development and through State highway planning programs whose funds are administered by the Department of Transportation.

Uses of the DIME File

This chapter describes the use of census and local data with the DIME file for planning and analysis in both the public and private sectors.

There are many potential uses for a DIME geographic base file or other geographic files incorporating DIME features. Several were explored by the Census Use Study and others await exploration. The description that follows is intended to provide a preliminary look at these uses.

Considerable research into uses of DIME files is being undertaken at the Southern California Regional Information Study (SCRIS), a joint project of the Census Bureau and the Southern California Association of Governments.[1]

Computer Mapping

Computer mapping using a DIME file can only be accomplished if the file contains coordinates.[2]

Node coordinates alone allow only one basic type of map to be produced—a network or outline map. These coordinates are used to plot street networks, to shade within the areas described by the nodes, or to print characters or symbols within the same area. Figure 6 illustrates an outline map with shaded node-defined areas.

To allow for full use of a DIME file for computer mapping, area centroid (center point) coordinates can be calculated and added to the DIME file. For example, block centroids are necessary to plot block data to a grid, as in figure 7; to a contour interval, as in figure 8; or to proximal or incidence map. Centroids are also necessary for mapping data to other areas, such as police precincts, census tracts, and transportation zones.

The Census Use Study tested several mapping techniques and systems using the DIME file, such as MAP 01, SYMAP, pen plotters, cathode ray tubes, and the Geospace plotter.[3]

In general there are six steps in computer mapping: Selecting and specifying the data to be mapped, selecting the mapping system to be used, attaching coordinates to the data by matching the data records to a DIME-type geographic base file with coordinates,[4] manipulating and organizing the data for input into the computer mapping system, selecting the appropriate cartographic features and data categories of the map to emphasize the salient characteristics of the data, and producing the map. Figure 9 illustrates this process.

---

[1] This research will be documented and reports released on a periodic basis.

[2] The Census Bureau presently plans to provide coordinate readings for all nodes in the geographic base files completed as part of the ACG/DIME Geographic Base File Program.

[3] For a full description of computer mapping research carried out at the Census Use Study, see Census Use Study report, *Computer Mapping.*

[4] This can be normally done by using an address matching system such as ADMATCH, described on page 18.

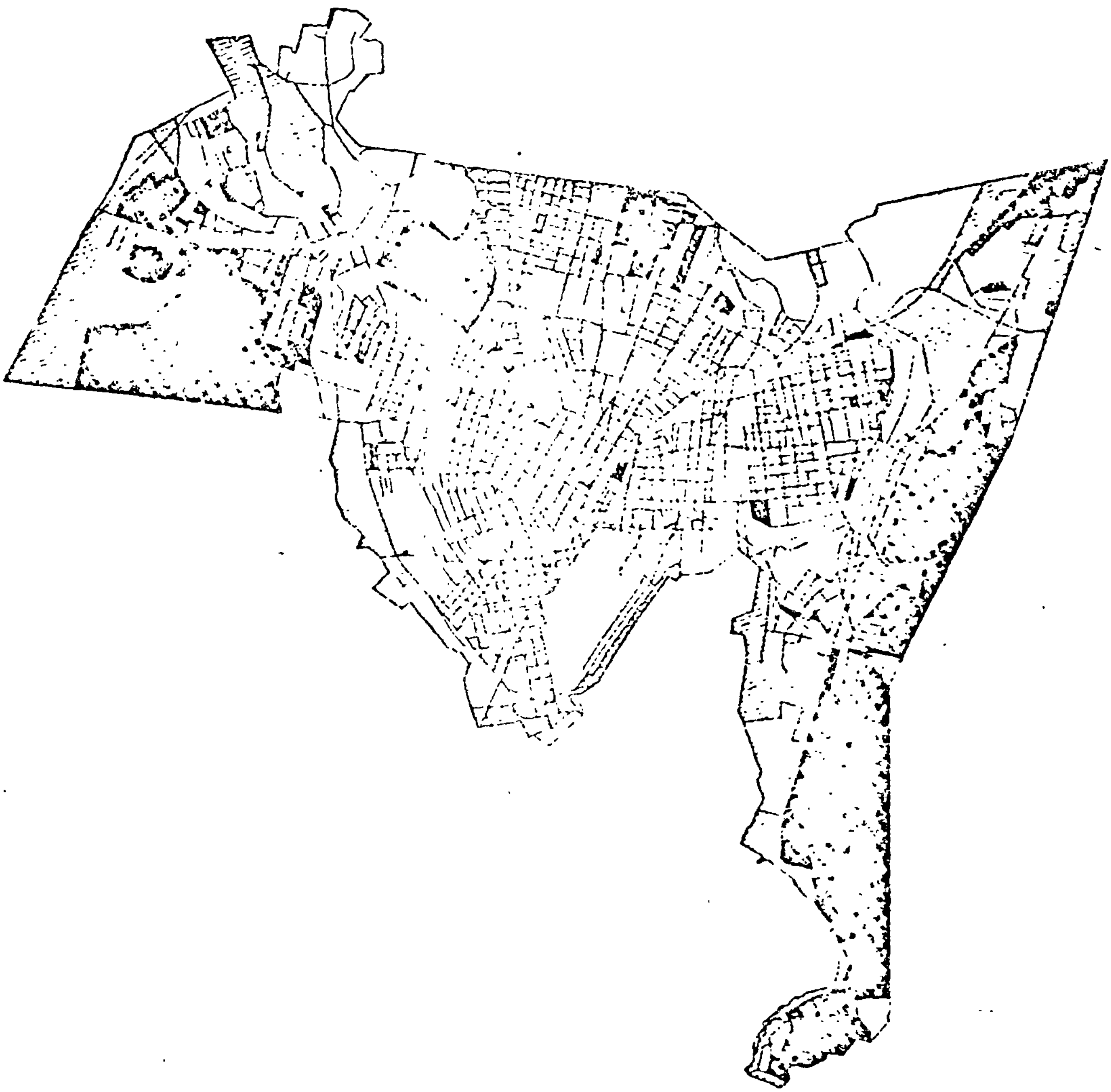


Figure 6. Outline Map—Geospace plotter shading map identifying blocks (percent of owner-occupied housing)

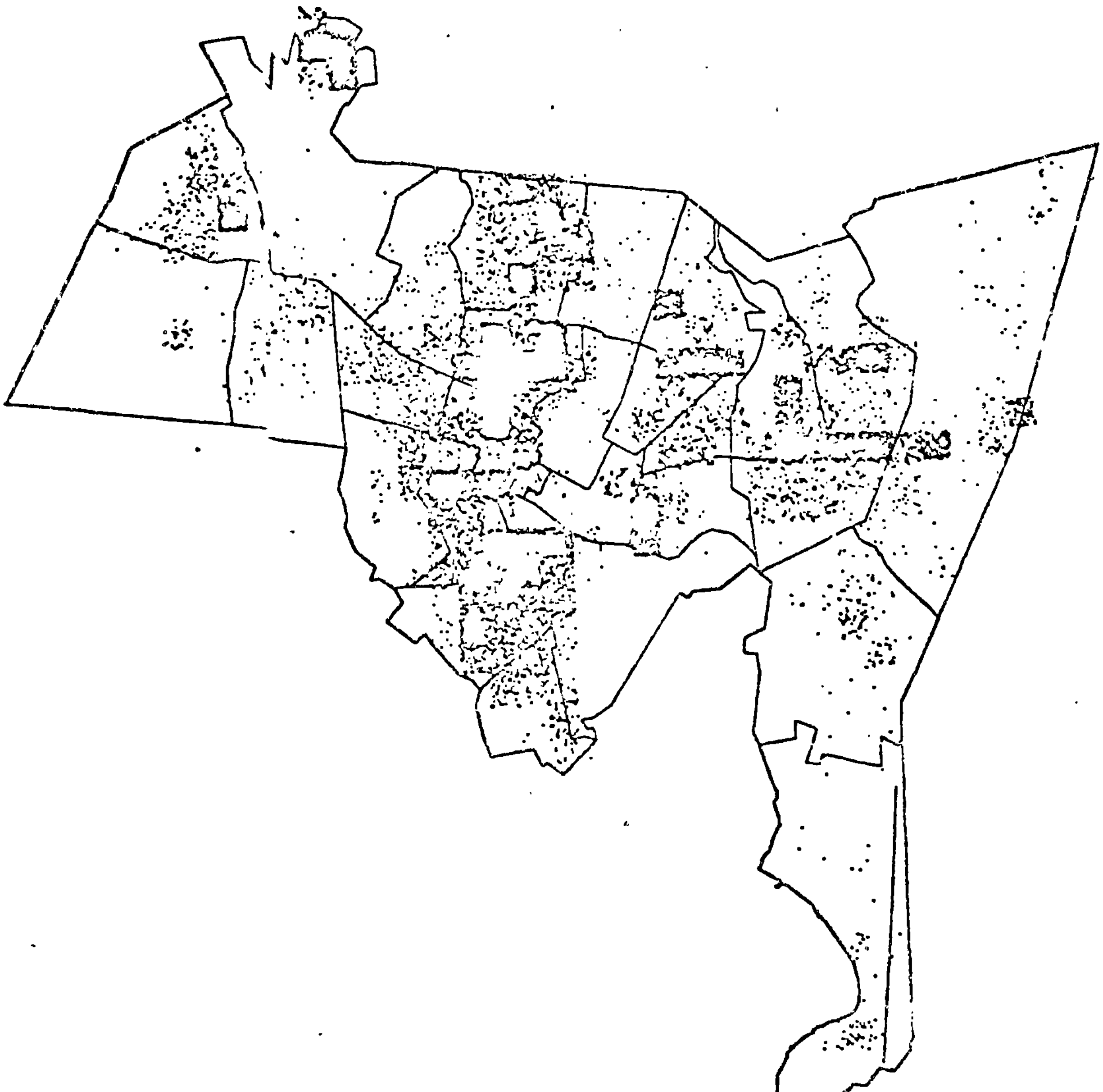


Figure 7. Grid Map Cathode ray tube map showing grid cells (adjusted hours spent by visiting nurses)

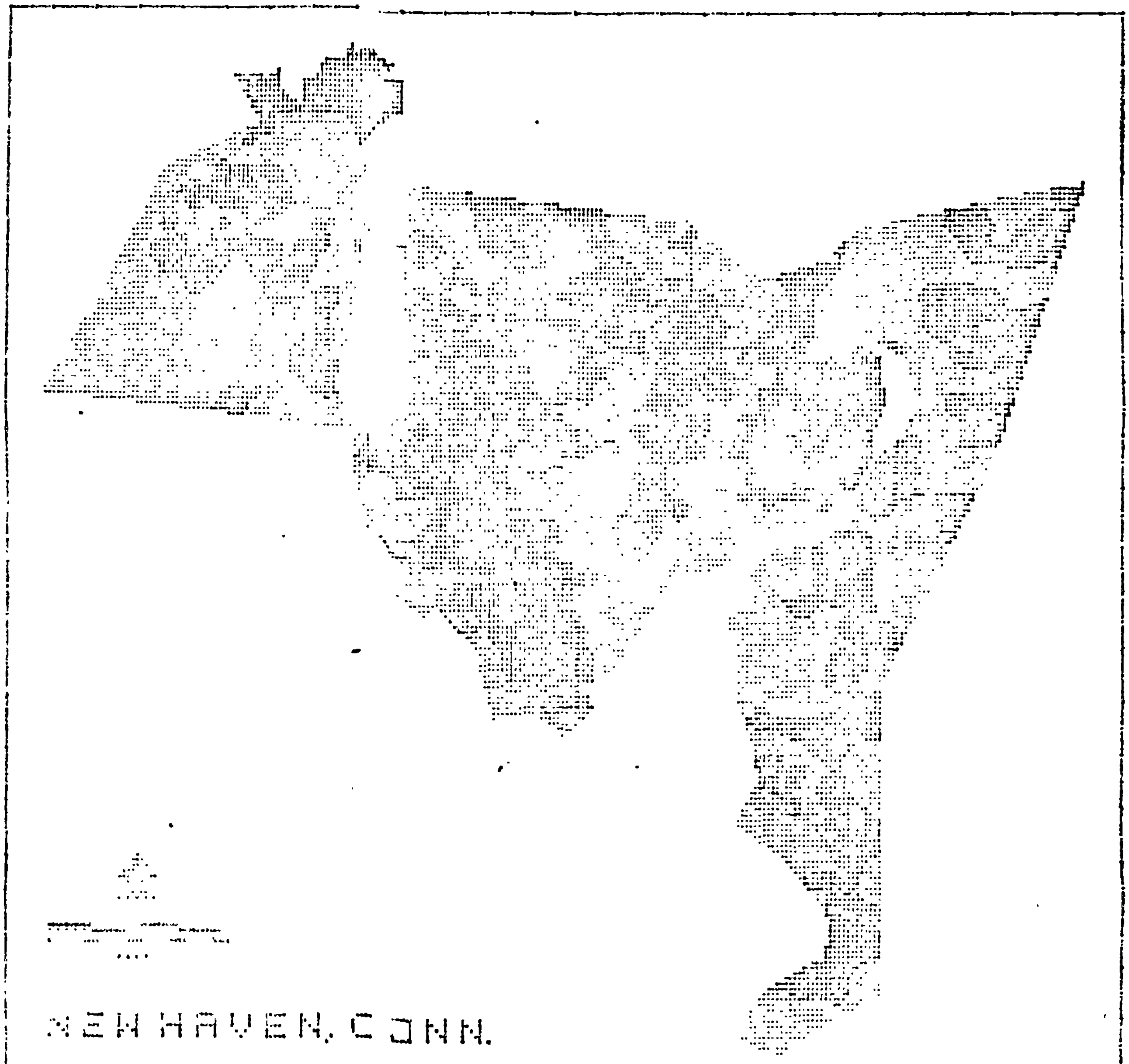


Figure 8. Contour Map—SYMAP showing contour levels (density of preschool children)

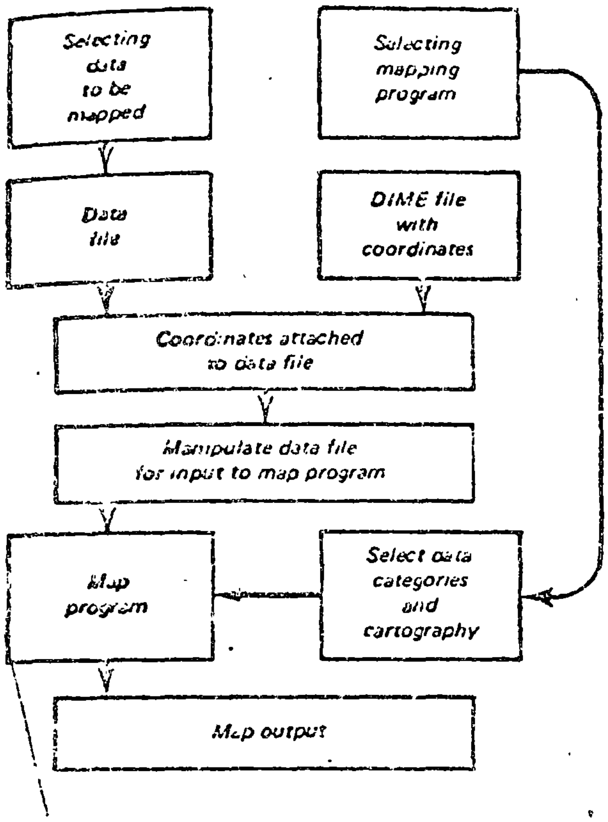


Figure 9. Computer mapping process

Adding Local Area Codes to the DIME File

The inclusion of local area codes allows use of the DIME file in the analysis of local and census data as they relate to local areas. Examples of local areas for which codes can be added to geographic base files include traffic zones, planning districts, school districts, police precincts, poverty neighborhoods, and health districts. Local area codes in a DIME file facilitate the interrelating of census and local data either for the areas originally coded in the DIME file, such as census tracts, blocks, municipalities, and ZIP code areas, or for local areas such as land use files or assessor records.

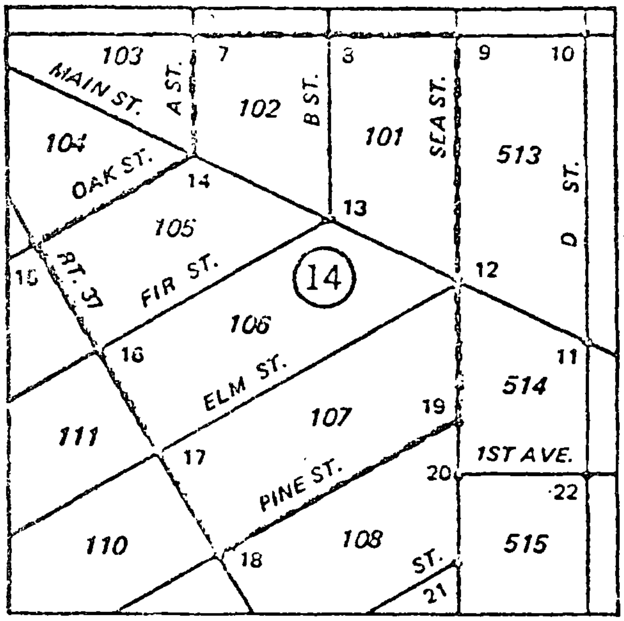
Local areas are usually defined as agglomerations of city blocks, parcels, or other small areas. However, a local area may also be defined in a linear sense, i.e., a series of street or other segments; or it may be identified by a series of points such as street intersections. The 1970 census data will be available only for areal units. The census block will generally be the smallest unit.[5] However, local data may be available for any of the three types of local areas.

Local codes may be added to a DIME file by means of a dictionary or correspondence table of local codes for each census area. There are two relatively simple methods which can be employed. The primary method is to plot the local areas on a Census Bureau metropolitan map and then manually prepare a corresponding list of local area codes for each individual or series of census area codes comparable to the local areas. The census codes necessary for areal units are census tract and block, as illustrated in figure 10. For linear coding of one segment at a time (streets or other segments) the census codes necessary are a combination of the census tract number and the node numbers at each end of the segment, as illustrated in figure 11. For linear coding of several continuous segments along a street, the necessary codes are the segment name and census tract and node numbers for the beginning and ending points of the contiguous segments, as illustrated in figure 12. For node or intersection coding, the necessary codes are census tract and node number, as illustrated in figure 13. Once a dictionary or correspondence table is prepared, it is keypunched and the file processed to add the local codes to the DIME records.

The variant method of creating an area dictionary utilizes the DIME file as the source for the dictionary records. A computer program is written to process the DIME file and punch a card for each unique occurrence of the basic component of the dictionary: node, segment, block, tract, etc. For example, if traffic zones are the local areal units to be included in the area dictionary, the tape file is processed and a card punched for each unique tract/block number combination. The cards are processed automatically in a card interpreter machine which imprints the cards with the coded elements so they can be read by clerks. These cards are manually divided into traffic zones, using a metropolitan map marked with traffic zone boundaries as a guide. The traffic zone identifiers are "gang-punched" into blank fields in the cards

---

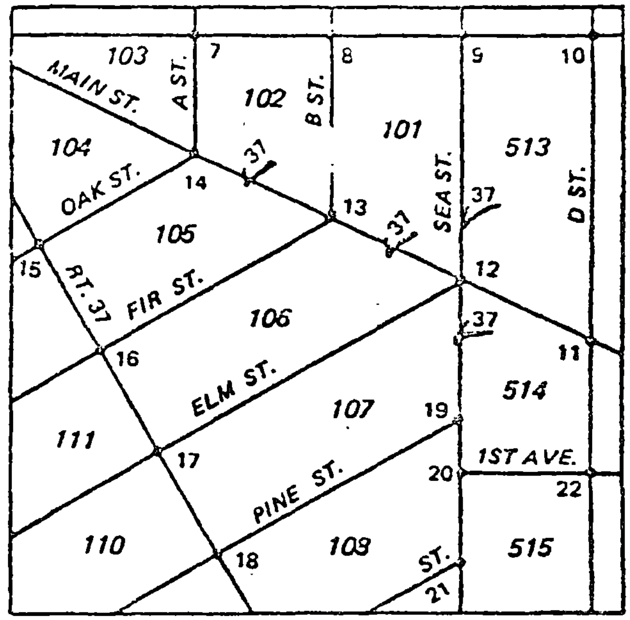
[5] Block data will only be available for Census Bureau-defined urbanized areas and other urban areas which contract with the Census Bureau for block statistics. An urbanized area contains a city (or twin cities) of 50,000 or more population plus the surrounding closely settled incorporated and unincorporated areas which meet certain criteria of population size or density. Block side data will be available for the urbanized area of many census SMSA's on a contract basis only.



Local area boundary
Local area code - (14)
Entire map in census tract 21.01 -

Local code	Census tract	Census block
14	21.01	101
14	21.01	102
14	21.01	105
14	21.01	106
14	21.01	107

Figure 10. Areal unit coding - local area 14 outlined on map coded to census tract and block.



Local linear code - 37
Coding key -
Entire map in census tract 11

Local code	From node	To node	Census tract
37	14	13	11
37	13	12	11
37	9	12	11
37	12	19	11

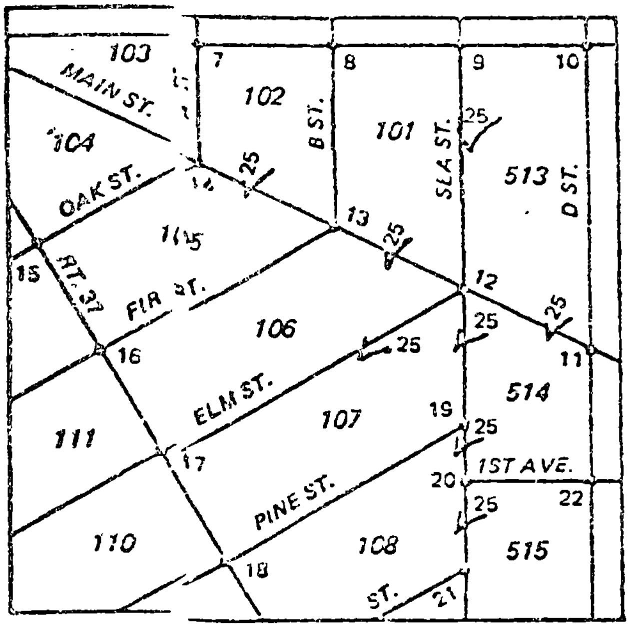
Figure 11. Linear coding by segment - local linear code /37 marked on the appropriate street segments of the map is coded to census tract, from node, and to node.

using a card reproducer machine. The card file is then usually converted to a tape file. Depending on the use which will be made of the file, it may or may not be combined with the original DIME file. When it is necessary to assign the local codes to the DIME file, the dictionary and the DIME file should be sorted by tract and block, and a program written to accomplish the code assignment.

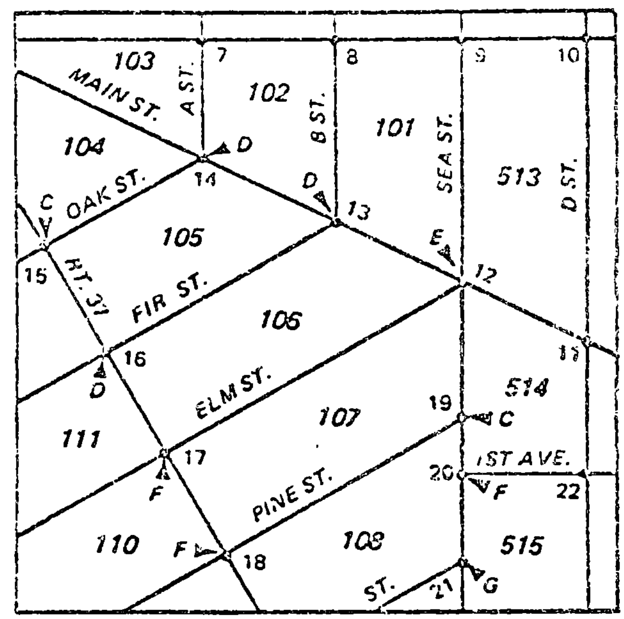
More sophisticated methods exist for adding local geographic codes to a geographic base file such as DIME. One example is the area bounding method, which describes the local geographic area by means of coordinates. For this method, a computer program could be written to determine the specific local area code for each record in the DIME file and insert the code into the record.

Using the dictionary method, the Census Use Study added local geographic area codes to the New Haven DIME file. These included codes for 1960 census enumeration districts, Community Action Program (CAP) areas and sub-CAP areas, elementary and secondary school districts, traffic zones, and telephone company central office and market areas.

Agencies requiring special local area tabulations were requested to provide a punchcard file consisting of one card per census block. The card contained the census tract and block and its corresponding local



Local linear code 25
Coding key - └─
Entire map in census tract 101



Local node code in italics - arrowed to node
Entire map in census tract 25

Local code	Street name	From node	To node	Census tract
25	ELM ST	17	12	*101*
25	MAIN ST	14	11	*101*
25	SEA ST	9	21	*101*

Figure 12. Linear coding by street - local linear code 25 marked on the appropriate street segments of the map is coded by street name, from node for the first segment along the street, to node for the last segment, and census tract.

Local code	Census tract	Node number
C	25	15
D	25	16
F	25	17
F	25	18
D	25	14

Figure 13. Node or intersection coding - local node codes (in italics) entered on the map and arrowed to the nodes they represent are coded by census tract and node number.

area code. If a block was partially in one local area and partially in another, the agency assigned the block to only one local area. This situation occurred infrequently because the local areas tended to be much larger than the average city block and to be bounded by streets or other features used by the Census Bureau to define blocks. For the telephone company central office and market area codes, however, a large number of arbitrary decisions had to be made because the areas were linear in nature, i.e., they were defined as being both sides of a street, rather than as a series of blocks.

When the punchcard file containing the local area code dictionary was received by the Census Use Study, it was computer edited to ascertain that each block was assigned to only one local area. After all local area block dictionaries were submitted and checked, one master computer tape containing all local area codes for each tract/block combination was created.

As in any large-scale data processing operation, certain problems were encountered due to clerical errors and omissions. The census data tape for New Haven city contained 993 unique city block codes. The DIME file and the Census Bureau metropolitan maps showed 1,117 blocks in the city. The difference between these numbers typically indicates the number of blocks with no population. However, it was discovered that 12 blocks appeared on the data tape with population figures; this error was due to clerical or processing errors in the construction of the original address coding guide for New Haven. These blocks did not appear in the local area block dictionaries. To add these blocks to the dictionaries it was necessary to refer to the address coding guide for street names and address ranges and then relate this information to the maps containing the local areas.

The most widespread use of local area codes was for creating tabulations of basic census data for the local areas. All special tabulations were checked before release to be sure of adherence to Census Bureau confidentiality rules. Special tabulations for any local area with fewer than five households were suppressed.[6]

Adding DIME File Codes to Local Records

Another valuable feature of a DIME-type file is its capability for address matching, the matching of records in two files on the basis of address numbers. When using a DIME file, address matching is the matching of local record files with address numbers to a DIME file with address number ranges to assign geographic codes to the local record files.

Address matching has many uses. For example, files containing records of building and demolition permits can be address matched to insert census tract numbers so that tabulations of new construction and demolitions can be prepared by census tract. These can be used with decennial census population data in estimating current population by tract. Address matching crime incidence records to census tracts or block groups can be used to indicate possible correlations between crime and census socioeconomic data. Address matching assessor records containing land-use data to planning district codes contained in the DIME file would show land-use patterns within the planning districts. Using the same assessor records matched to geographic coordinates in DIME file, computer maps of the land-use patterns can be prepared.

To facilitate address matching, the Census Use Study has developed an address matching computer package named ADMATCH.[7] The ADMATCH system can be used with DIME files, address coding guides, ACG/DIME geographic base files, census tract street indexes, or any similar computer file containing addresses or address ranges and geographic area codes. The ADMATCH system requires that data and reference files (DIME, ACG, etc.) be processed in a three-phase operation:

1. Preprocessing—The ADMATCH Preprocessor program deciphers the address and creates a 75-character match key. Special features of the Preprocessor include the ability to create city and State codes, standardize street types (e.g. ST, AVE), correct misspelled street names on data records, and add records to the reference file to compensate for street names that are misspelled or truncated.

2. Sorting—The data and reference files must be sorted on certain items in the match key, including ZIP code or State and place (city) codes, street name, and house number or address range.

3. Matching—The ADMATCH Matcher program compares items in the match keys of the data file and reference file. When a match occurs, user-specified geographic codes from the reference file are attached to the data records. Since the quality of data files varies, the Matcher program has an option to accept mismatches on the street type and direction to improve the chances for matching. For example, the Matcher can be instructed to accept a record such as 151 Elm St. when matched to 151 Elm, or 76 W. Washington St. N.W. when matched to 76 Washington St. N.W. Individual street addresses can be matched against street addresses as well as address ranges.

The concept of the matching process is illustrated in figure 14.

In New Haven, several local files were matched to the DIME file to produce tabulations and input to computer mapping programs. Local files matched included birth records, police complaints, police arrests, building code violations, fires, school attendance records, school census records, visiting nurses hours, and health records. The match rate averaged 85 to 90 percent for a perfect match (all items match perfectly) and it approached 100 percent for less than a perfect match (some items allowed to mismatch).

In one experiment, data on visiting time spent by nurses at various locations throughout New Haven

[6] For details, see Census Use Study report, *Data Tabulation Activities*.

[7] See Census Use Study computer program package, *ADMATCH: An Address Matching System*.

were provided by the Visiting Nurses Association of New Haven. This file, consisting of 29,000 records, was matched to census tract and block and coordinates in the New Haven DIME file. A 98.6 percent match rate resulted. Once matched, the local data record file with coordinates became input to a computer mapping program. Maps displaying the visit locations were prepared. Tabulations of the data were also prepared for census tracts and blocks.

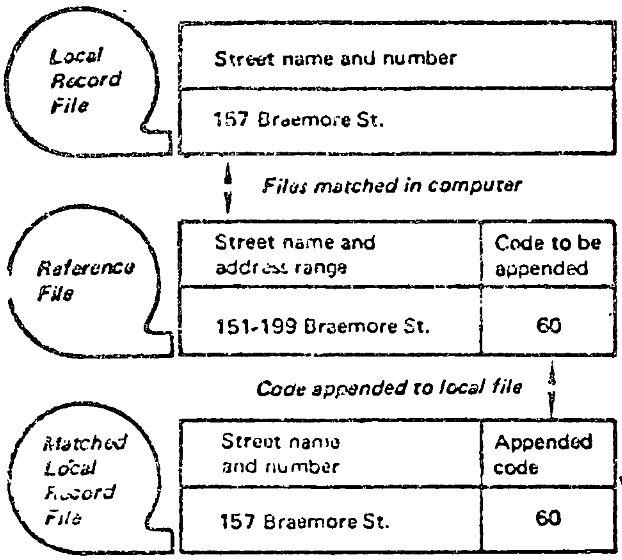


Figure 14. Matching process.

ADMATCH was also used extensively in matching local records for input to the Health Information System.[8] Among the items matched were birth records, infant and fetal death records, and hospital obstetrical records. These local record files were matched to the DIME file, and census tract and block group codes appended. Once matched, the local records were tabulated, analyzed, and input into the Health Information System.

Network and Node Analysis

Nodes or link-node combinations are used in node analyses; calculation of area, density, and network analyses; and calculation of centroids for blocks, segments, or block sides.

Node Analysis

A DIME file or other geographic base file with node coordinates can be used to geographically order data by street intersections, such as data on traffic accidents at intersections. These data can then be compared to information on traffic flow and the existence of traffic signals and signs and pedestrian crosswalks at intersections with high accident rates. Analyses can be made concerning changes needed in speed limits, the timing of traffic lights, the need for warning signs, additional pedestrian crosswalks, stop signs, etc.

Processing is accomplished in a number of ways. Local data can be coded directly to node numbers in a DIME file by clerical transcription and keypunching methods. Or a separate intersection file can be created by sorting the DIME file on node number and creating separate records for each intersection (containing geographic identifiers, node numbers, and intersecting street names—listed once for each node). This file can be sorted on intersecting pairs of streets in standard alphabetic order and local data records for intersections can be added.

A special matching program that can correct and match on misspelled street names or add supplementary records could also be created to match local intersection data to the DIME intersection file and transcribe geographic or other codes from one file to the other. Figure 15 illustrates the concept of matching a DIME intersection file and a local intersection file and appending a code from the DIME file to the local file.

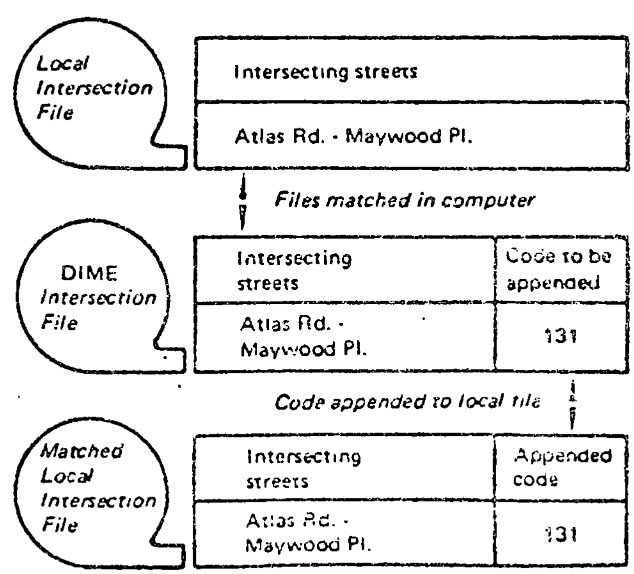


Figure 15. Matching intersection files

[8] For a full explanation, see Census Use Study report, *Health Information System*

Area Calculations and Arrays

Polygonal areas can be computed given coordinate information describing the location. Polygons can represent agglomerations of blocks, census tracts, police precincts, or arbitrary areas (centered on a set of nodes or one node) of fixed or variable width. For example, an area calculation may be required for a strip 100 feet wide on both sides of a series of street segments, as one of the data elements needed in a study of land requirements for highway construction. Figure 16 illustrates this. Or an area calculation may be required for an area, radiating out ¼ mile from a set of nodes, describing a particular set of conditions such as a parcel of land proposal for rezoning (figure 17). Or again, area calculations may be required for rings and sectors at fixed distances from a central business district.

Also, given data by polygons described in terms of coordinates, it is possible to determine by "point-in-polygon" or "polygon-in-polygon" methods which polygon a given data item belongs to. This permits arraying data by irregular polygons.

Similar but simpler methods permit arraying data by regular polygons such as grid squares, triangles, circles, rings, or sectors.

Network Analysis

A DIME file forms a substantially complete link-node system or network. Networks are used in transportation planning and resource allocation studies in which movements take place from one point to another over a network.

Transportation Studies—An important device in transportation studies is the simulation of traffic flow over a network using minimum path algorithms. This device can be used for processes such as assignment of delivery trucks and postmen to least time/distance paths to make routing more efficient; assignment on a real-time basis of police cars, ambulances, fire trucks, etc., from one point in a network to another over a least time/distance path; estimations of the effects on traffic patterns when elements of the network change; analysis of effects of changes in traffic generators, etc.

Planners have used network files representing major traffic arteries for many years in conducting

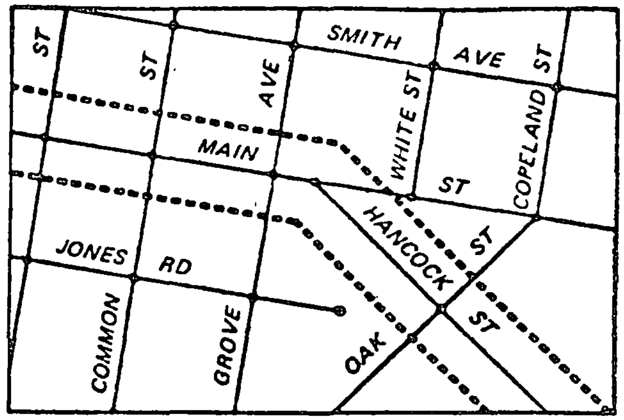


Figure 16. Area calculation for a strip centered on a series of linear segments.

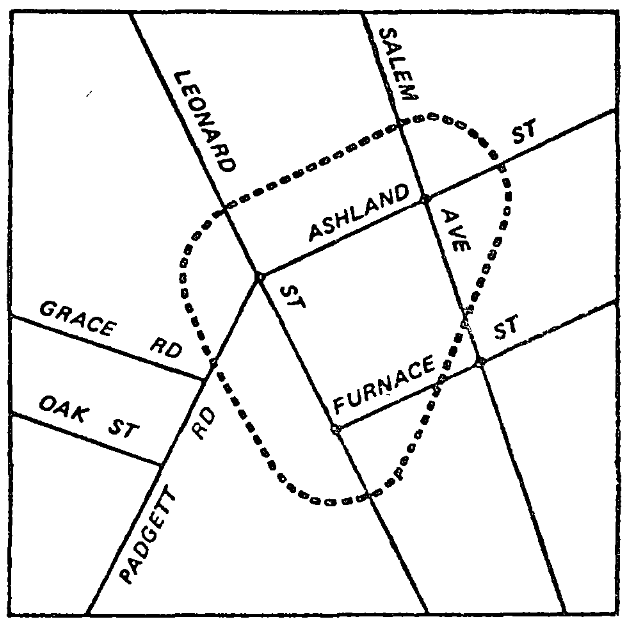


Figure 17. Area calculation for an area radiating out a certain distance from a set of nodes.

transportation studies. A DIME file contains all streets and many nonstreet features and boundaries for an area. In some cases, it may contain too many records to be handled by computer programs currently used in transportation planning. A DIME file may also lack data important to transportation system simulations, such as direction for one-way streets, number of vehicle lanes in a street, speeds, and travel time.

A DIME file can be readily adjusted for transportation applications. Data needed can be selected by developing a dictionary of required streets, comparing this to the DIME file, and creating a new network file for only the required streets. It is also possible to add data on direction, capacity, etc., to a DIME file. Figure 18 illustrates this process.

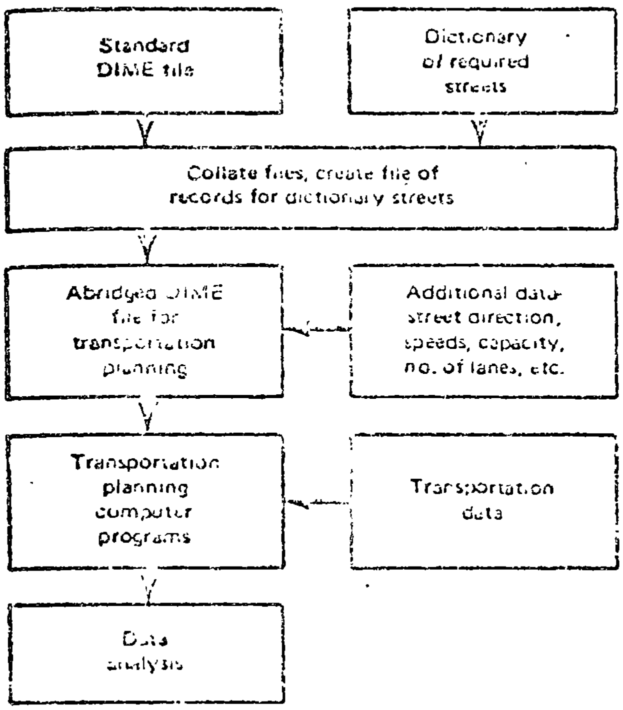


Figure 18. Adapting the DIME file for transportation planning.

Transportation network data currently being used vary considerably in format, content, coverage, and level of detail from one area to another. Standard DIME files could be used as the basis for development of standard transportation network data throughout the country. Standard computer programs, technical specifications and analytical documents would result in considerable savings and increased efficiency. Standard DIME files for all major metropolitan areas would make it worthwhile to develop standard computer programs for solving allocation problems involving delivery truck routes, trash removal routes, bus routes, computer routing of vehicles, "dial-a-bus" systems, analysis of potential loading on existing and proposed mass transportation routes, etc.

Allocation of Resources to Facilities—With the advent of the DIME file, the Office of Civil Defense realized that substantial benefits could be gained from standardizing the Community Shelter Planning (CSP) process throughout the country. The Census Use Study is designing and developing a computerized model which will allocate people to community fallout shelters.

In the past, the CSP process relied on local information and manual map-oriented data manipulation to locate shelters, identify population distributions by small areas and by time of day, and assign people to shelters to minimize the distance traveled and use available shelter spaces most efficiently.

The 1970 census block data will minimize the need for detailed local data. Uniform DIME files for all the links in the total transportation system of a metropolitan area will make it feasible to assign people to shelters by computer. To modernize the CSP process, two objectives are being pursued in the Census Use Study project:

1. Development of a computerized model to automatically assign people to fallout shelters along a DIME network, and

2. Preparation of a data management system capable of assembling and organizing census and local data into the form required by the model.

A preliminary computer model has been developed—CRAM (Computer Resource Allocation Model). It represents the first use of the DIME file as a network. The spatial relationships between shelter facilities and people using the shelters (population data by census blocks) are determined through the use of the DIME file. The model can perform an allocation based on pedestrian and/or vehicular modes of travel. Figure 19 illustrates the CRAM model.

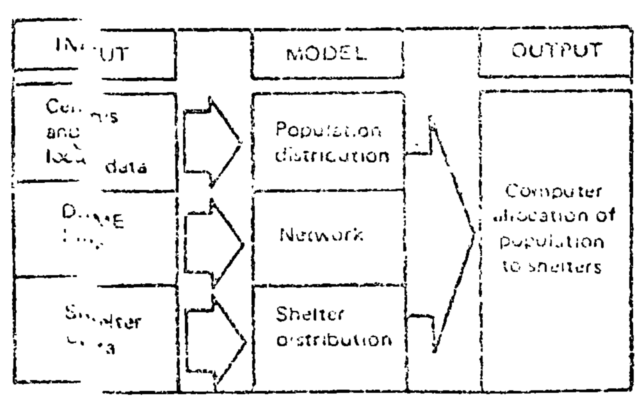


Figure 19. Network allocation of population to shelters (CRAM) model.

Although the model has been developed to assign people to shelters, it can be used for a wide range of allocation problems—for example to allocate children to schools, determine logical service areas for community health facilities, evaluate alternative sites for new plants, evaluate existing facilities, and plan more effective utilization of existing facilities. Since the model allocates resources to facilities along a transportation network, it may be applied to any situation which embodies these three components.

Adjacency Analysis

The ability to group contiguous blocks (or other small geographic areas) by certain homogeneous characteristics of the blocks has interested urban planners and analysts for some time. A DIME file and appropriate software make this process faster and more efficient by relating all nodes, segments, and blocks to each other. Just as network analysis uses the link-node relational characteristics contained in a DIME file, adjacency analysis uses block-node relational characteristics.

Aggregation of Areas

Contiguous areas such as blocks, tracts, and police beats can be combined to create an area of predetermined size in terms of land area, number of households, number of street segments, etc. If the areas are to be defined in terms such as square feet or acres, a DIME file with coordinates would be the only required input to the computer process for aggregation. If the areas are defined in terms of other data, a data file containing the necessary data elements must also be used.

Computer programs for aggregation can be designed to provide compact aggregations so that fingers, holes, or serpentine aggregations are avoided. This can be done by specifying in the computer program that each successive area added to the aggregation should be the area closest to the initial area. Divergencies from this rule can be allowed. For example, a 40-percent divergency would allow one dimension of the aggregation to be 40 percent greater than the other. Other constraints such as a feature to stay within block groupings or census tracts as much as possible or to stay within municipal boundaries, can be added.

A prime use for aggregation techniques would be the delineation of legislative districts, based on the one man-one vote ruling of the Supreme Court. Although the results of an aggregation program may not be completely acceptable to lawmakers responsible for drawing such districts, they would provide a solid base or starting point from which equal population districts could be delineated. Other uses would be the devising of police patrol beats based on the accumulated length of the segments within a compact aggregated area, or the determination of school or health districts based on existing and/or proposed schools or clinics.

Clustering of Similar Areas

Contiguous areas, all within the same range of data values, can be accumulated into larger areas of undetermined size using a DIME file, data files, and appropriate computer software. Data values are the basis for the clustering.

The computer software may be designed so data value ranges can be stipulated prior to operating the computer program or so that they can be determined mathematically during computer processing in terms of quartiles, quintiles, etc. Also, the software may provide for stipulating size criteria for clusters (e.g., maximum or minimum number of areas) as well as criteria on data value discontinuities.

The concept of clustering allows for the determination of homogeneous areas for community action programs, model cities development, poverty areas, high health risk areas, ethnic enclaves, areas where the elderly predominate, or other types of homogeneous areas.

Other Uses

Other potential uses for a DIME file include—

Monitoring of Programs by Geographic Areas

A DIME file framework can be used for reporting the detailed analysis data on the progress of a program for a specific geographic area, ranging from the block level upwards. The Model Cities Administration of the Department of Housing and Urban Development is currently exploring the possibility of using DIME files for this purpose.

Development of Parcel Files

The DIME techniques for creating a geographic base file can be used to create a parcel file. Coding and processing would probably be expensive. The limited amount of research carried out by the Census Use Study to date indicates that a DIME-type parcel file would cost at least 20 times more than a typical DIME file at the block level. The methodology is identical: Blocks become parcels; nodes are placed at parcel corners; and each lot or parcel receives a unique number, like a block number. Address ranges are coded for each parcel, if available. Since a geographic base file must contain records for all the geography within an area, sidewalks, streets, intersections, and other elements such as traffic islands, cul-de-sacs, bodies of water, and bridges, should be considered as parcels and so coded.

For some applications, it would not be necessary to create a parcel file in the detail outlined above. For each segment record in a regular DIME file, subsidiary records for each parcel along the segment could be created. Using this method, parcel data could be assigned through the segment record to the parcel subsidiary record. Parcel data could then be retrieved at the parcel, segment, block, or any higher geographic level.

Development of Larger-Grained Files

Just as it is possible to create DIME files for parcels, it is also possible to create files for areas larger than blocks such as block groups (or enumeration districts),[9] census tracts, planning districts, police beats, municipalities, and counties. For example, a researcher concerned with rural problems might wish to create a DIME file for a large depressed rural area. Such a file could be based on enumeration districts, these being the smallest tabulation units available from a decennial census for rural areas. The researcher would then be able to manipulate census data, and local data coded to the ED, using the techniques outlined previously in this chapter.

---

[9] The geographical unit called the block group corresponds to the enumeration district in previous censuses. There are approximately 10 blocks per block group. The number of block groups per census tract varies from one to eight; the average is four. The block group is always a subdivision of a census tract.

Voting ward or precinct DIME files could be used by political scientists in analyzing election returns or by politicians designing new legislative district boundaries. County DIME files could be used by economic planners at the State or Federal level to study and analyze economic trends and projections.

Information Systems

The DIME file can be readily used as the geographic base for information systems. For example, in the Health Information System[10] developed in New Haven by the Census Use Study for the State of Connecticut, block groups derived from the DIME file were the basic geographic units used in the system. Local data files were matched to block groups by means of the ADMATCH system. Local and census data were combined at the block group level and various statistical analyses were performed to generate socioeconomic and health indicators.

Information systems for municipalities and other political or planning jurisdictions should be based on a geographic framework such as DIME. By using a DIME file, the data elements in an information system could be manipulated, analyzed, and displayed using network and node analysis, adjacency analysis, computer mapping techniques, etc.

Area Sampling

Random or stratified geographic sampling on a segment side, segment, block, or higher geographic level could be accomplished using a standard DIME file. Sampling at the parcel or large-grained level could also be done if such files were created.

Geographic Grouping of Census Summary Tape Data

The use and manipulation of data provided on a block and block grouping basis by the Census Bureau after the 1970 census will be greatly facilitated by the use of DIME files.

---

[10] For further information see Census Use Study report, *Health Information System*. While this report details a system primarily concerned with health, its methodology is widely applicable to other functional areas.

Time Series Studies

With a geographic base file such as DIME, data can be coded to the block, segment, or other level to reflect historical developments for time series analyses.

Spatial Comparisons

Researchers may wish to compare and study differences in the characteristics of certain areas from one city to another. Using some of the techniques described in this chapter, such study areas could be delimited and various data items retrieved and studied for the areas. As an example, socioeconomic characteristics could be obtained for a series of hard-core unemployment areas in various cities throughout the country from 1970 census summary tapes and analyzed using the appropriate DIME files and analytical techniques.

Market Analysis and Other Business Uses

Many of the techniques described above have applications in market analysis. For example, market analysis areas can be inserted in a DIME file, resulting in the ability to retrieve census or company data by market areas. Network and node analysis has value in the determination of delivery routes. Use of a DIME file with lists of prospective customers gives a firm the ability to map out a rational scheme for following up on such prospects. Placement of new facilities or relocation of existing facilities because of changes in the nature of business or customer characteristics can be planned by using the appropriate software, data, and DIME files.

Technical Description of the DIME System

This chapter describes the conceptual origins and technical aspects of the DIME system.

Conceptual Origins

The most significant technical contribution of the DIME geocoding system is the topological edit. It provides for accurate computer editing and correcting of the structural elements of the coded geographic file; i.e., streets and other linear elements, points where the linear elements intersect, and area identifiers. The edit can theoretically be done manually, although practically it is only done by computer, except in a demonstration or test situation. Depending on the number of times the file is cycled through the edit process, it is possible to correct the clerically coded file so that it becomes a perfect replica of the elements on the map coded. However, in practice, cost limitations usually do not permit a sufficient number of edit cycles to correct the file completely. A residue of uncorrected errors may range from less than 1 percent to 4 or 5 percent. At least two or three edit cycles should be completed to assure sufficient accuracy in the file.

As stated earlier, the edit system is based on concepts derived from graph theory. Since any street map is basically a type of linear graph, concepts derived from graph theory can be used as a means for creating, correcting, and maintaining a high quality geographic base file. Single-line maps can be categorized as simple geometric figures consisting of three basic uniquely identifiable elements. Points or vertices, lines connecting these points, and areas enclosed within a series of lines forming an enclosed space. Figure 20 illustrates these elements of a typical single-line map.

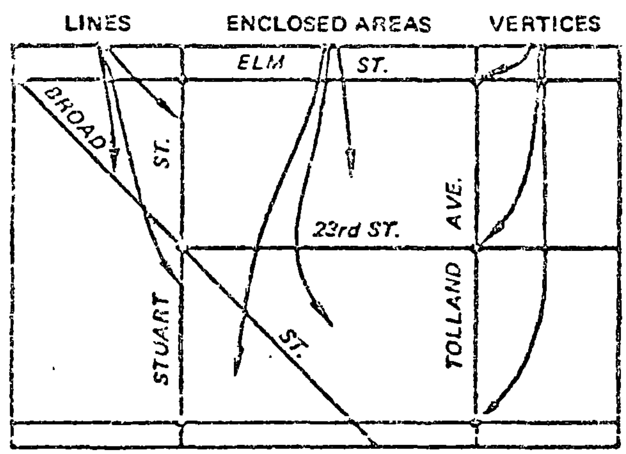


Figure 20. Linear graph elements of a map.

The Census Bureau's series of metropolitan maps is ideally suited for this purpose because all three elements can be identified uniquely. Two of these elements are identified on the maps when produced: lines are identified as Grant St., Muddy River, etc., and areas are identified by block numbers, census tract numbers, etc. The third element can be identified by uniquely numbering all vertices or points where lines begin or end. Curved streets or other linear features on the map, such as rivers, railroad tracks, or boundaries, can be represented as a series of straight-line segments by defining a sufficient number of vertices along the curved line.

Composition of DIME File

A DIME file is composed of segment records. A segment is defined as a length of a street or other feature between two distinct vertices or nodes. Other features are imaginary lines defining political or other boundaries; topological features such as rivers, shorelines, and canals; other map features such as railroad tracks, airport runways, and piers; and any other feature defining a block boundary. Nodes are points where features begin, end, intersect, or curve sharply. Unnamed features, when coded, are described uniquely.

Each segment is coded separately with the three basic codes needed to edit the file—segment name describing the linear element, "from" and "to" node numbers describing the nodes at either end of the segment, and left and right block numbers describing the areal identifiers on either side of the segment. If the segment is a street, address ranges for both sides of the segment are usually coded. Address ranges are coded because DIME files will frequently be used to assign geographic codes to local data files containing street addresses as the only locational code. There is also a separate coding field which can be used to describe nonstreet features and other unique segments such as proposed streets, pedestrian walkways, lake shores, rivers, and boundaries.

Additional code fields which are used for a series of segments are ZIP code, area code (a municipality code), election ward numbers, and census tract. It is usually not necessary to code ZIP codes or election ward numbers if they will not be used. If more than one municipality is coded, area codes should identify each segment to its proper municipality. If the area has census tracts, they should be coded. If census tracts do not exist for an area and the size of the area is small enough so that each block within the area can be uniquely identified with a three-digit code, then census tract codes need not be used. "Pseudo census tract" areas can be created and numbered for nontract areas where an areal identifier larger than a block but smaller than a municipality is needed.[1]

In summary, the elements in the DIME file are:

Mandatory elements
- Segment name and/or description
- Node numbers (two for each segment)
- Block codes (two for each segment)

Elements mandatory under certain conditions
- Address ranges
- Area codes (municipality code)
- Census tract codes

Optional elements
- Code for nonstreet features
- ZIP code
- Ward or other election district codes

Figures 21 and 22 illustrate a listing of a typical DIME file.

DIME Topological Edits

There are two types of DIME edits. One "chains" the string of segments that bound a block and is therefore known as the block chaining edit. The other "chains" the blocks surrounding a node and is known as the node chaining edit. The block chaining edit is performed first and is the most important edit as it detects the great majority of structural defects in the coded file. Because of this, the node chaining edit is frequently omitted.

Block Chaining Edit

The block chaining edit operates on the three mandatory coded elements for each segment record: Segment name or description, node numbers, and block codes. It also serves as a check on the accuracy of the census tract code as the records are sorted by tract prior to the computer processing. As a byproduct of the topological edits, erroneous tract codes are detected. An elementary illustration of the method used in the block chaining edit is shown below. Block 105, the block to be edited, is shown in figure 23. The basic elements needed for the edit of block 105 are shown in figure 24.

The basic steps followed by the computer are:

1. All segments coded to block 105 (either block-left or block-right) for the census tract being edited are selected from the file.

2. As each segment record for block 105 is selected, the computer checks the position of the block number of the block being edited.

---

[1] A full explanation of the coding methods, supervision, problems, and alternatives is provided in the clerical instructions manual of the Census Use Study computer program package, *DIME: A Geographic Base File System.*

	[illegible] Code		Ward		Census tract		Header No.
	Left	Right	Left	Right	Left	Right	
11001	35	35	7	7	14	15	30151

Figure 21. Header items (codes for a series of segments)

Segment name or description	Code	From node	To node	Block No.		Left Addresses		Right Addresses		Header No.
				Left	Right	Low	High	Low	High	
ANDERSON RD.		75	76	111	120	900	998	901	999	30151
ANDERSON RD.		76	77	112	119	1000	1098	1001	1099	30151
ANDERSON RD.		77	78	113	118	1100	1198	1101	1199	30151
ANDERSON RD.		78	79	114	117	1200	1248	1201	1249	30151
ARGONNE ST.		34	36	271	279	400	488	401	449	30151
ARGONNE ST.		36	35	270	283	450	498	451	499	30151
ARGONNE ST.		35	39	270	282	500	598	501	599	30151
BADGER RIVER	2	107	108	137	137					30151
BADGER RIVER	2	108	112	137	137					30151
BADGER RIVER	2	112	113	138	137					30151

Figure 22. Segment items (codes for each segment)

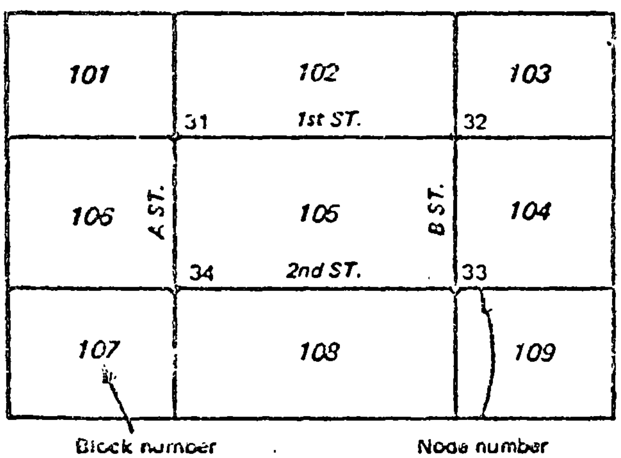


Figure 23.

Segment name	From node	To node	Block left	Block right
1st St.	31	32	102	105
2nd St.	34	33	105	108
A St.	34	31	106	105
B St.	33	32	105	104

Figure 24.

a. If the block number is in the block-left position, it is transferred to the block-right position and the other block number is transferred to the block-left position. The node numbers are also exchanged; the "from" node replaces the "to" node and vice versa.

b. If the block number is in the block-right position, no changes are made.

3. When all the block numbers for the block being edited are in the block-right position, the computer attempts to link or chain the nodes from one record to another, rearranging the sequence of segments as necessary. Notice that it was necessary to move the last segment record in figure 25 to a position between the first and second records. Figure 26 illustrates the final arrangement of the segments and the dotted lines indicate how the computer chains the segment records.

If the nodes chain and the first "from" node is the same as the last "to" node the block is considered topologically correct. Note the parallel of the computer operation in the hypothetical chaining of block 105 in figure 27.

Segment name	From node	To node	Block left	Block right
1st St.	31	32	*102*	105
2nd St.	33	34	*108*	105
A St.	34	31	*106*	105
B St.	32	33	*104*	105

Figure 25.

Segment name	From node	To node	Block left	Block right
1st St.	31	32	*102*	105
B St.	32	33	*104*	105
2nd St.	33	34	*108*	105
A St.	34	31	*106*	105

Figure 26.

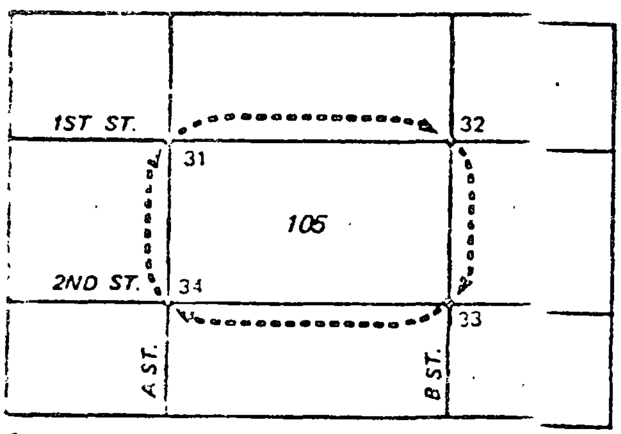


Figure 27.

If any segments remain, or if the block cannot be chained, the block records are rejected as a potential error. For instance, if any of the records in the above example were missing (i.e., not coded) the block would not chain and would therefore be rejected.

If the node numbers or block numbers were reversed, the block would not chain properly and would be rejected. As an example, if the left and right block numbers for 1st Street in figure 24, were coded 105 to the left rather than to the right and 102 to the right rather than the left, the block would contain a "reversal" and would be rejected. Figure 28 illustrates this point.

Segment name	From node	To node	Block left	Block right
B St.	32 →	33	*104*	*105*
2nd St.	33 ↙ →	34	*108*	*105*
A St.	34 ↙ →	31	*106*	*105*
1st St.	32	31	*102*	*105*

Figure 28.

Segment records for the blocks rejected are printed out on a reject listing for review. When reviewed, and corrected or recoded, the segment records are keypunched, inserted in the computer file, and reedited.

Node Chaining Edit

The node chaining edit operates with the same elements, and chains blocks around a node rather than nodes around a block. An elementary example of the method used is illustrated below. Node 29, the node to be edited, is shown in the center of figure 29. The basic coded elements needed for the edit of node 29 are shown in figure 30.

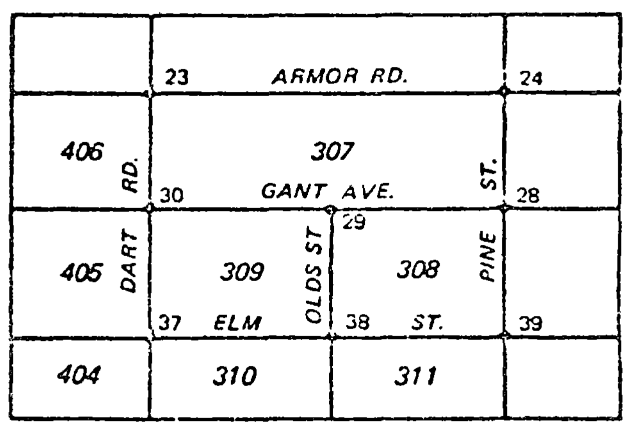


Figure 29.

Segment name	From node	To node	Block left	Block right
Gant Ave.	30	29	*307*	*309*
Gant Ave.	29	28	*307*	*308*
Olds St.	38	29	*309*	*308*

Figure 30.

Essentially the same process is used in the node chaining edit as is used in the block chaining edit. The basic steps are:

1. All segments coded to node 29 (either in the "from" or "to" node position) for census tract being edited are selected from the coded file.

2. As each record is selected the node number for the node being edited is automatically transferred (if necessary) to the "to" node position as illustrated in figure 31. If the node number is transferred the block numbers are also exchanged.

Segment name	From node	To node	Block left	Block right
t Ave.	30	29	307	309
Gant Ave.	23	29	306	307
Olds St.	38	29	309	308

Figure 31.

3. When the segment records are structured so that the node being edited is in the "to" node position, an attempt is made to chain the blocks around the node, rearranging the sequence of segments as necessary. Notice that it was necessary to move the last segment record in figure 31 to a position between the first and second records. Figure 32 illustrates the final arrangement and the dotted lines indicate the chain around the node.

Segment name	From node	To node	Block left	Block right
Gant Ave.	30	29	307	→ 309
Olds St.	38	29	309	→ 308
Gant Ave.	28	29	308	→ 307

Figure 32.

If the blocks chain and the first left block is the same as the last right block, the node is considered to be topologically correct.

If any segments remain, or if the the node cannot be chained, the node records are rejected as a potential error. For example, if any of the records in the above example were missing, i.e., not coded, the node would not chain and would, therefore, be rejected. If the node numbers or block numbers were erroneously reversed, the node would also be rejected.

All segment records for the nodes rejected are printed out on a standard reject listing for review. When reviewed, and corrected or recoded, the segment records are keypunched, inserted in the computer file and reedited.

Thus, two basic structural elements in the file can be made 100-percent accurate by processing the file until all rejects are corrected. The other structural element—segment name (and address numbers)—is edited during the address edit, discussed below

Interior Segment Edit

There is a limitation to the topological edits. Segments interior to a block such as dead-end streets cannot be edited. For example, figure 33 illustrates a dead-end street--Pine Place, which does not form part of the boundary of block 305. When edited, the segment for Pine Place would be rejected as a potential error because it does not form part of the chain around block 305. When investigated, it would be found that Pine Place is a legitimate segment record and as a result would be retained in the file as correct. However, if Pine Place were missed in coding, it would not have been detected by the topological edit because the block would have been chained and found acceptable.

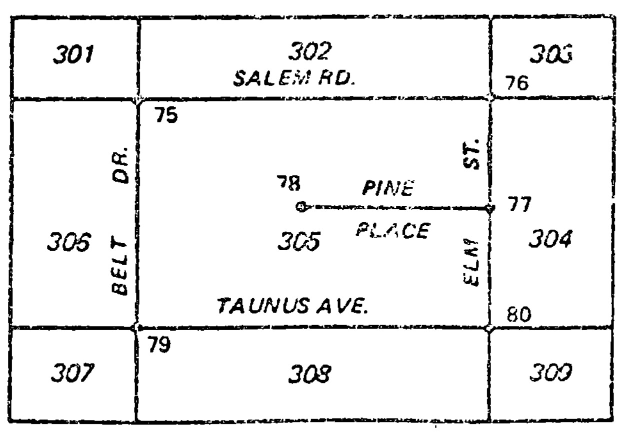


Figure 33.

Segment name	From node	To node	Block left	Block right
Salem Rd.	75	76	302	305
Elm St.	76	77	304	305
Elm St.	77	80	304	305
Taunus Ave.	80	79	303	305
Bels Dr.	79	75	300	305
Pine Place	77	78	302	305

Figure 34.

There is a relatively simple method of assuring the inclusion in the file of all interior segments. The method entails matching node numbers in the file against a listing of nodes developed clerically when the coding maps are node numbered. The development of this listing on a node control form is discussed further in the clerical manual of the DIME computer program package. The node control listing includes inclusive ranges of all node numbers used in the area coded. Therefore, each of the node numbers listed on the node control form should match at least once to the node numbers in the DIME file.

Matching can be done either manually or by computer. To accomplish a manual match, each node in the DIME file should be computer listed once—in map number, census tract number, node number order. The resulting listing should then be matched to the node control list, which is prepared in the same order. All unmatched node numbers should be investigated for possible errors. The computer matching process requires the development of a computer program to compare the two lists in the order indicated above and print out all unmatched cases. The node control form should be keypunched for insertion into the program. Unmatched cases should be investigated for possible errors.

DIME Address Edit

The DIME address edit was developed primarily to check the completeness and consistency of address ranges and street names in a DIME file. The edit checks one entire street at a time by stringing together all segments for the street by linking node numbers. Once linked, the address ranges on each side of the street must be in ascending order without overlaps. One side of the street must have even address numbers and the other, odd address numbers. If a street is broken into pieces or passes from one jurisdiction to another, appropriate exceptions are allowed. Nonstreet records are not edited.

Coordinates

The node numbered maps used in coding can be digitized; i.e., coordinates can be determined for each node, at any time after coding has been completed. However, the insertion of digitized coordinates into a DIME file is not usually performed until the file is computer edited as it is desirable to have a "clean" file for coordinate insertion.

The primary reason for adding coordinates to a DIME file is to prepare the file for computer mapping, distance calculations, and other applications involving spatial relationships.

There are a number of different coordinate systems and each has its own strengths and weaknesses. The Census Use Study used only the state plane coordinate system. As mentioned earlier in this report, the Census Bureau plans to make the ACG/DIME geographic base files available with geographic coordinates (latitude-longitude), in state plane coordinates and also with "map miles" north and east from an arbitrary point.

Several methods can be used to digitize a map and insert coordinates in a DIME file. The method used by the Census Bureau entails the use of a semiautomatic coordinate locator built by Bureau technicians. Generally there are five steps in the process. The first step is to set up the map on the coordinate locator (digitizer) table and prepare the map for digitizing. The second step is to read or digitize the map coordinates for each node. The third step is to convert the map coordinates, which are read in hundredths of an inch by the digitizer, to actual ground coordinates—state plane, geographic (latitude-longitude) coordinates, etc. Some digitizing equipment combines steps two and three. The fourth step is to attach the coordinate records to the DIME file records to which they apply. The fifth step is to plot the DIME file with coordinates at the same scale as the original map, compare the two maps, and correct any obvious errors. These five steps are explained in more detail below.

Map Setup

The map sheet to be digitized is placed on the digitizer table and positioned so that the state plane coordinate registration marks are parallel to the table edges. The registration marks, located on the edges of the maps, are aligned parallel to the table edge because in some cases map edges may not be parallel to the coordinate system indicated along those edges.

After each map is set up, an origin point is determined. This point should be to the left (west) of and lower than (south of) the most southwesternly node to be digitized. This point should be located in the border region of the map and clearly marked and identified on the map. The origin point can be any arbitrary point meeting the above requirements. The digitizer should then be adjusted to measure from the origin point.

Locations of at least two registration marks on each border are read and the coordinates and digitizer readings for these registration marks recorded for later use by the conversion programs.

Digitizing

Digitizing is the process of reading map coordinates for each node from the node-numbered coding map. In using the Census Bureau's semi-automatic coordinate locator, the operator of the machine places a cursor over each numbered node and presses a button which records or reads the location of the node. The node is read in hundredths of an inch up and to the right of an origin point located in the lower left corner of the digitizing machine. Each time a node is read, the operator must also keypunch the map sheet number, census tract number, and node number on a punchcard, thus creating a complete record for each digitized node.

Some digitizing equipment have the ability to automatically multiply each coordinate reading (in hundredths of an inch) by the appropriate scale factor and add the resulting reading to the geographic (latitude-longitude) or state plane coordinate reading of the origin point.

For small areas it is often practical to digitize the map manually on a drafting table. This is done by selecting an origin point on the map to the left and below the lowest left node to be digitized, and then measuring carefully the location of each node relative to the origin point. As each node is read, a record of its X and Y location, node number, and any other necessary map code should be prepared for keypunching. A variant on this method would be to use a light-table, and place graph paper under the map to be digitized. The grid lines on the graph paper should be systematically numbered according to some pre-defined scaling method. The map is placed over the graph paper and the nodes read by relating the node to a pair of grid lines, scanning the grid lines and recording the appropriate X-Y readings. Another method would be to transcribe the map onto graph paper and proceed as outlined above.

However the digitizing is done, certain quality control measures should be followed. It is less costly to catch errors at the initial digitizing stage than after coordinates are inserted in the DIME file. A suggested procedure is as follows:

1. After all nodes on a map have been read in the normal manner by one operator, a second operator rereads every 25th node starting at a randomly selected node in the first 25. The rereading must be independent, without reference or comparison to the original readings.

2. A third person compares the readings of the original operator to the readings of the second operator and notes all differences. Differences of less than .05 inches can be ignored. Each difference is checked to ascertain which operator made the error. If the original operator had more than a certain prespecified percentage of node readings in error (5 to 10 percent, depending on accuracy desired), the original digitizing is rejected. The map is then completely redigitized. Then, a new quality control sample is used. If less than the prespecified percentage of node readings are in error, the work is accepted and the node readings in error are corrected.

After all incorrect readings are corrected, the digitizing process is complete. During digitizing and quality control, the map should not be demounted from the digitizer because reading errors may result if the map is remounted and digitized. Quality control procedures should always be used, regardless of the reputation of the operator or the organization.

For digitizing at the Census Bureau, the node numbers to be digitized on each map are first selected from the ACG/DIME file and punched out on cards. This is done to eliminate manual entry of node numbers and is made possible by the fact that the Census Bureau Coordinate Locator can display prepunched information from the cards on the operator's console.

Conversion

This is the process of transforming the original digitizer readings (in hundredths of an inch) to a coordinate system. Conversion is usually done by multiplying each node reading in hundredths of an inch by the appropriate map scale and adding the result to the coordinate reading of the origin point. A simple computer program can be written to perform this conversion.

Scale factors and the state plane coordinate reading of the origin point are determined as follows:

1. Divide the difference between all pairs of state plane coordinate readings along each border of the map by the same difference in inches. The scales should be nearly identical for all borders. Significant differences should be checked, small differences ignored. Paper maps tend to have greater differences than more stable material such as Mylar..

2. Record in inches the readings of the registration marks around the map border. Multiply each registration mark reading by the scale factor and subtract the product from the state plane coordinate reading for the registration mark. The average of the numbers resulting from this subtraction yields the state plane coordinate reading for the origin point.

Similar processing is applied to convert digitizer readings to other coordinate systems.

Coordinate Insertion

Coordinate insertion requires that another computer program be written. The program reads and stores the coordinatè file in a directly accessible medium, either in its entirety or in parts such as in map sheet or tract number parts. The DIME file (either in its entirety or in map sheet or tract number parts) is then processed, one segment record at a time. The coordinate file is searched for the coordinate reading for each end (node) of the DIME segment record. Node coordinates are then appended to each DIME record.

This program can also be written to perform a rudimentary edit by printing out a listing of all segment records for which coordinates are not available and all coordinate records which were not appended to a DIME record at least once. The program can also check segment records which seem to be excessively long for the nature of the area being digitized.

A subsequent program may be used to perform certain mathematical checks for errors of closure and the existence of intersections between segments which do not share a node. Such intersections would be evidence of digitizing errors since all legitimate intersections should be nodes.

Plotting Test Map

Once the final DIME file with coordinates is prepared, it is plotted with line plotter equipment at the same scale as the original map. The resulting map should be carefully compared to the original map and any serious divergencies noted. If there are sufficient numbers of serious reading errors to warrant correction, an interpreted punchcard file of the segment records containing coordinates is prepared. Each reading error is corrected by measuring the X and Y distances between the incorrect reading and its correct location, converting the distances to map scale, and entering the correct readings on the appropriate punchcards. These punchcards are then repunched and merged with the original file to create a final corrected DIME file with coordinates.

Development of a DIME File

This chapter describes in general terms the process of creating a DIME file from clerical coding through computer processing. It also describes coordinate insertion, and updating and maintenance research.

DIME or ACG/DIME geographic base files will be available beginning in 1970 for the urbanized areas in most of the 233 standard metropolitan statistical areas of the country. Regional planning agencies or councils of government will often be the agencies responsible for development of the file. In other areas, county planning agencies or transportation studies may have this responsibility.

For the remaining urban areas of the country, including those SMSA's which did not opt for the Census Bureau's ACG/DIME program and smaller cities not eligible for the program, local agencies, with adequate technical staff, can create a DIME file. Otherwise, a computer consultant or software firm can be hired to assist in creating a DIME file using the clerical instructions and computer procedures and programs in the Census Use Study computer program package, *DIME: A Geographic Base File System.*

Other research and planning organizations who wish to create a DIME file for experimentation may also obtain copies of the computer program package.

Figures 35, 36, 37, and 38 outline the preparation of maps for coding, clerical coding, computer processing, and insertion of coordinates.

Prerequisites

Creation of a DIME file, requires consideration of costs, technical and coding staff, computer requirements, coding maps, and address reference materials.

Costs

Preparation of a DIME file without coordinates by local agency personnel alone should cost approximately $2.00 to $2.50 per block for clerical coding and computer processing, assuming accurate and suitable scale coding maps, address reference materials, and the proper computer equipment are available. The digitizing and insertion of coordinates into the file should add $.50 to $1.00 per block, making a total of $2.50 to $3.50 per block.

Technical and Coding Staff

For most coding operations five coders and one supervisor will probably be sufficient. With five coders a 1,000 block area would take approximately 1 month to complete—1 to 2 days training, 1 to 2 days node numbering, 5 to 7 days coding, 1 to 2 days clerical edit and quality control, and 3 to 4 days each for clerical correction of the topological and address edit rejects. If the coding area is considerably larger than 1,000 blocks, it might be desirable to increase the coding staff to 10 or 15, keeping constant the 1 to 5 supervisor-coder ratio.

One professional should be responsible for the entire operation and the person should be familiar with the geography of the area and with computer processing capabilities and techniques. Specific suggestions concerning the coding staff are contained in the clerical instructions of the DIME computer package.

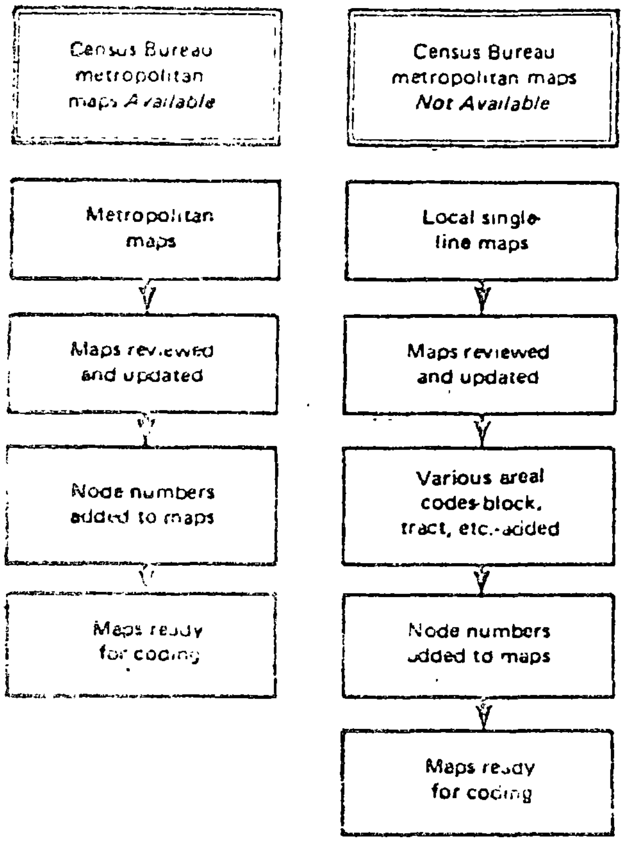


Figure 35. Preparation of maps for coding.

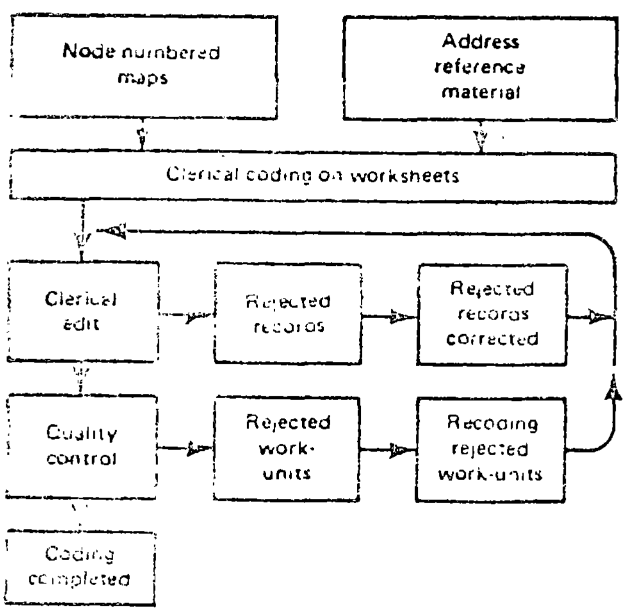


Figure 36. Clerical coding of DIME file.

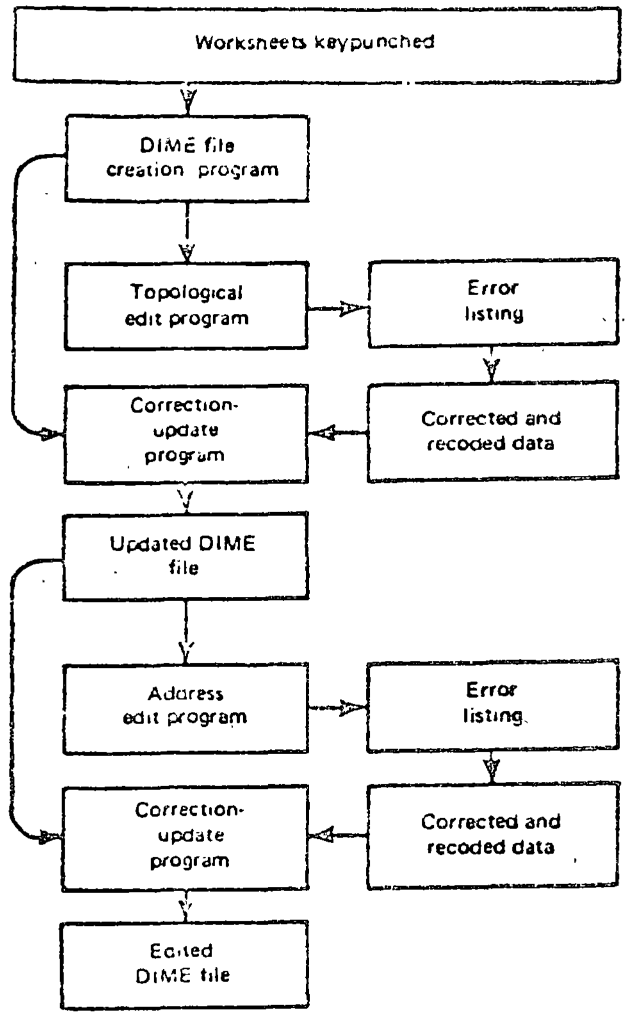


Figure 37. Computer processing of a DIME file.

Computer Requirements

DIME computer programs were written to run on IBM System 360/40 disk operating system with core capacity of 65K. The following devices are also required: A standard keypunch machine for key-punching the coding worksheets, a card reader for transmitting punched card data input, and a tape drive for the DIME master file. The DIME file may be held on direct-access (disk) storage, in which case sufficient space must be available. Direct-access storage, preferably on disk, is also needed to hold intermediate files. All programs are written in G-level FORTRAN IV.

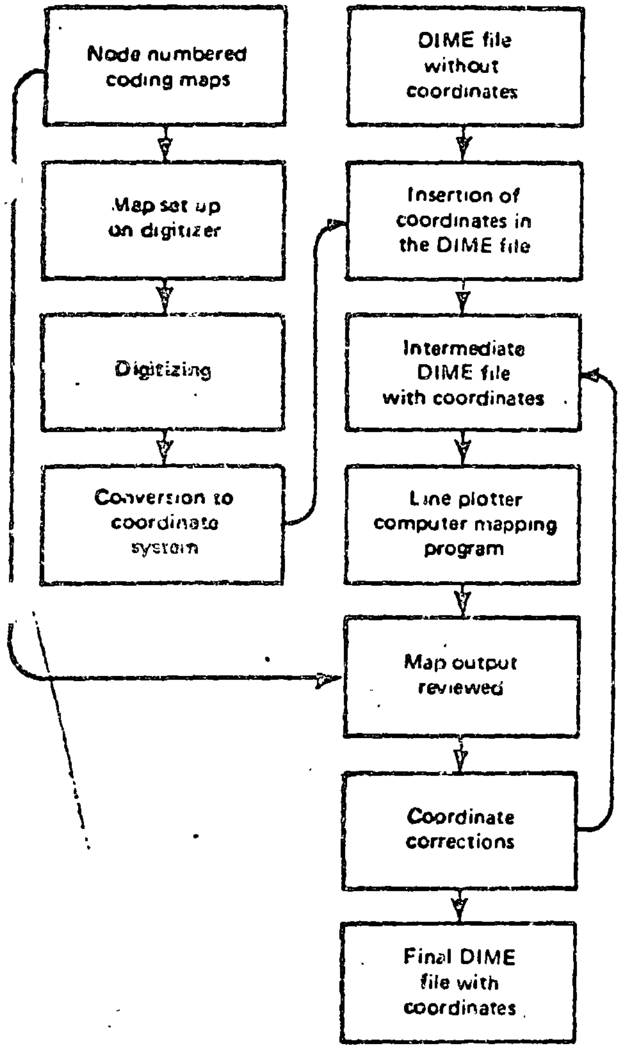


Figure 38. Coordinate digitization and insertion process.

The DIME computer package tapes are available in industry compatible seven- and nine-channel tape format. Documentation and technical specifications for the DIME computer edits and processing can be found in the computer manual of the DIME computer package.

Coding Maps

The requirements for coding maps are discussed fully in the clerical instructions of the DIME computer package.

Generally the scale of the coding maps should be within the range of 1 inch = 400 feet to 1 inch = 1,000 feet. Single-line maps are strongly recommended. They should contain all existing streets and street names, municipal boundaries, railroad tracks, and drainage features such as lakes and rivers.

Address Reference Materials

The address reference materials used in coding must contain the following information: Street names, intersecting streets (and, if possible, other intersecting features), and address numbers or ranges between intersects. The even-odd address number dichotomy should be apparent. Usually address reference materials will be maps, but they may also be street address listings, street or city directories, or some other type of index. All address reference material should be field checked for accuracy before using. If no accurate reference material can be located, the field listing form and instructions provided in the clerical instructions should be followed.

Clerical Coding

The clerical instruction manual contained in the DIME package describes the various steps and processes necessary to complete the coding phase of DIME file creation. It contains chapters on personnel and space requirements, materials to be used in coding, address reference materials, coding maps, the preparation of coding maps for coding, the preparation of special instructions for local problems, assignment preparation, training, and supervision of coding and postcoding operations. It also contains appendixes with sample forms, field listing instructions, and a coder's manual containing complete coding instructions.

The chapter on supervision of coding operations contains sections dealing with making assignments, controlling the operational flow, checking coders' work, and recordkeeping. It also provides technical procedures for coding regular and irregular or arbitrary address ranges and systems; problems such as unnamed streets, unknown street names, proposed streets, nonvehicular streets, addressable features other than streets, and the adding and deleting of node numbers.

Computer Processing

The computer procedures manual of the DIME package describes the computer processing steps required in a DIME operation. It also describes the overall system design and the hardware and software environment.

Each of the four computer programs comprising the system are described in full. These four programs are (1) the master file creation program, (2) the topological edit program, (3) the address edit program, and (4) the correction/update program. The section on master file creation discusses input requirements program operation diagnostics, and output. The topological edit section includes a description of the input, editing procedures, output, and recoding. The address edit section includes input requirements, editing procedures, and output. The correction/update program section includes input requirements, diagnostics, and output.

Also included are descriptions of the sort/merge program, job control language, and sample listings.

Coordinate Insertion

Coordinates can be measured and recorded from the node-numbered coding maps by using a digitizer or other coordinate reader and they can then be converted to geographic coordinates and inserted into the DIME file.

For those areas of the country in the Census Bureau's ACG/DIME program, coordinates will be added to the file by the Bureau as an integral part of the program. Coordinates in this program will be latitude longitude carried to four decimal places for all nodes in the file. There are also plans to convert the latitude-longitude coordinates to state plane coordinates as an option.

Agencies that want to create their own DIME file will have to contract for digitizing services. Information on the technical aspects of coordinates is provided in chapter III.

The Census Use Study has had limited experience with organizations that perform this service. However, it is estimated that the cost of digitizing will run from 10 to 15 cents per node; or for an area of 1,000 blocks at 2.5 nodes per block, S250 to S375. These costs do not include conversion of the coordinates or insertion of the coordinates into the DIME file, a cost which could equal the digitizing cost.

File Maintenance

The Bureau plans to maintain the ACG/DIME geographic base files, probably updating them yearly. However, no operating system for updating and maintaining the files has yet been devised. One of the primary responsibilities of the Southern California Regional Information Study (SCRIS) is research into and development of operational methods for the periodic updating and maintenance of the interrelated ACG/DIME geographic base files and metropolitan mapping series.

If the Census Bureau receives approval from the Congress for a quinquennial census in 1975, updating of the geographic base files and the maps must begin in 1972 or 1973. Greater mail coverage in the census will require expansion of the coverage of the files and maps. If the ACG/DIME geographic base files are used for area sampling after 1970, it will perhaps be necessary for the Bureau to update the files on a more frequent basis. In any case, a technique and procedural requirements will be developed not only for use by the Bureau but also for use by local agencies who desire to update the files and maps periodically. The Bureau may provide a complete package for updating, maintenance, and expansion of the files. As research at SCRIS and the Census Bureau progresses, reports will be prepared describing the techniques and systems for file maintenance and updating.

Availability of Census Bureau Geographic Files

The alphabetical listing below includes all 233 standard metropolitan statistical areas. Listing anomalies are explained in the footnotes. There may be future additions to or deletions from this listing.

Address coding guides (ACG's) are now available, at least in preliminary form, for all mail census SMSA's. Final ACG's for all SMSA's will become available during the first quarter of 1970. They will available on either seven- or nine-track magnetic tape or on high-speed printer output. As an example of costs, the preliminary ACG for the Fort Wayne, Ind. SMSA, with an estimated 1966 population of 234,000, costs $36 for a seven-track and $42 for a nine-track tape. Another example, the Milwaukee, Wis. SMSA, with an estimated 1966 population of 1,335,000, costs $52 for a seven-track and $74 for a nine-track tape. Further information and order forms can be obtained from Central Users Service, Bureau of the Census, Washington, D.C. 20233.

The ACG/DIME or ACG Improvement Program geographic base files for mail census SMSA's will not begin to become available until late 1970. The SMSA's listed below are participating in the program or have expressed their intent to participate in the program.

ACG/DIME files for nonmail census SMSA's will begin to become available in mid-1970. Further information can be obtained from the Central Users Service, Bureau of the Census.

SMSA	Mail census ACG	Mail census ACG/DIME	Nonmail census ACG/DIME
Abilene, Tex.			X
Akron, Ohio	X	X	
Albany, Ga.			X
Albany-Schenectady-Troy, N.Y.	X	([1])	
Albuquerque, N. Mex.			X
Allentown-Bethlehem-Easton, Pa.-N.J.	X	X	
Altoona, Pa.	X	X	
Amarillo, Tex.			X
Anaheim-Santa Ana-Garden Grove, Calif.	X	X	
Anderson, Ind.	X	([1])	
Ann Arbor, Mich.	X	X	
Asheville, N.C.			X
Atlanta, Ga.	X	X	
Atlantic City, N.J.	X	([1])	
Augusta, Ga.-S.C.			X
Austin, Tex.			X
Bakersfield, Calif.			X
Baltimore, Md.	X	X	
Baton Rouge, La.			X
Bay City, Mich.	X	X	
Beaumont-Port Arthur-Orange, Tex.	X	X	
Billings, Mont.			X
Biloxi-Gulfport, Miss.			([2])
Binghamton, N.Y.-Pa.			X
Birmingham, Ala.	X	X	

See footnotes at end of list.

SMSA	Mail census ACG	Mail census ACG/DIME	Nonmail census ACG/DIME
Bloomington-Normal, Ill.	X	([1])	
Boise City, Idaho			X
Boston, Mass.	X	X	
Bridgeport, Conn.	X	([1])	
Brockton, Mass.	X	X	
Brownsville-Harlingen-San Benito, Tex.			X
Buffalo, N.Y.	X	X	
Canton, Ohio	X	X	
Cedar Rapids, Iowa			X
Champaign-Urbana, Ill.	X	([1])	
Charleston, S.C.			X
Charleston, W. Va.			X
Charlotte, N.C.	X	X	
Chattanooga, Tenn.-Ga.			X
Chicago, Ill.	X	X	
Cincinnati, Ohio-Ky.-Ind.	X	X	
Cleveland, Ohio	X	X	
Colorado Springs, Colo.			X
Columbia, S.C.			X
Columbus, Ga.-Ala.			X
Columbus, Ohio	X	X	
Corpus Christi, Tex.			X
Dallas, Tex	X	X	
Davenport-Rock Island-Moline, Iowa-Ill.	X	X	
Dayton, Ohio	X	X	
Decatur, Ill.	X	([1])	
Denver, Colo.	X	X	
Des Moines, Iowa			X
Detroit, Mich.	X	X	
Dubuque, Iowa			X
Duluth-Superior, Minn.-Wis.			X
Durham, N.C.	X	X	
El Paso, Tex.[3]	X	X	
Erie, Pa.			X
Eugene, Oreg.	X	X	
Evansville, Ind.-Ky.			X
Fall River, Mass.-R.I.	X	X	
Fargo-Moorhead, N. Dak.-Minn.			X
Fayetteville, N.C.			X
Fitchburg-Leominster, Mass.	X	X	
Flint, Mich.	X	X	
Fort Lauderdale-Hollywood, Fla.	X	X	
Fort Smith, Ark.-Okla.			X
Fort Wayne, Ind.	X	X	
Fort Worth, Tex.	X	X	
Fresno, Calif.			X
Gadsden, Ala.			X
Galveston-Texas City, Tex.	X	X	
Gary-Hammond-East Chicago, Ind.	X	X	
Grand Rapids, Mich.	X	([1])	
Great Falls, Mont.			X
Green Bay, Wis.	X	X	
Greensboro-Winston-Salem-High Point, N.C.	X	([1])	
Greenville, S C.			X
Hamilton-Middletown, Ohio	X	X	

See footnotes at end of list.

SMSA	Mail census ACG	Mail census ACG/DIME	Nonmail census ACG/DIME
Harrisburg, Pa.	X	(1)	
Hartford, Conn.	X	X	
Honolulu, Hawaii			X
Houston, Tex.	X	X	
Huntington-Ashland, W. Va.-Ky.-Ohio			X
Huntsville, Ala.			X
Indianapolis, Ind.	X	X	
Jackson, Mich.	X	(1)	
Jackson, Miss.			X
Jacksonville, Fla.	X	X	
Jersey City, N.J.	X	X	
Johnstown, Pa.	X	X	
Kalamazoo, Mich.	X	X	
Kansas City, Mo.-Kans.	X	X	
Kenosha, Wis.	X	X	
Knoxville, Tenn.			X
Lafayette, La.			(2)
Lafayette-West Lafayette, Ind.			(2)
Lake Charles, La.			(2)
Lancaster, Pa.	X	(1)	
Lansing, Mich.	X	X	
Laredo, Tex.			X
Las Vegas, Nev.			X
Lawrence-Haverhill, Mass.-N.H.	X	X	
Lawton, Okla.			X
Lewiston-Auburn, Maine			X
Lexington, Ky.			X
Lima, Ohio	X	X	
Lincoln, Nebr.			X
Little Rock-North Little Rock, Ark.			X
Lorain-Elyria, Ohio	X	X	
Los Angeles-Long Beach, Calif.	X	X	
Louisville, Ky.-Ind.	X	X	
Lowell, Mass	X	X	
Lubbock, Tex.			X
Lynchburg, Va.			X
Macon, Ga.			X
Madison, Wis	X	X	
Manchester, N.H.			X
Mansfield, Ohio	X	X	
Mayaguez, P.R.			(4)
McAllen-Pharr-Edinburg, Tex.			X
Memphis, Tenn.-Ark.	X	X	
Meriden, Conn.	X	(1)	
Miami, Fla.	X	X	
Midland, Tex.			X
Milwaukee, Wis.	X	X	
Minneapolis-St. Paul, Minn.	X	X	
Mobile, Ala.	X	X	
Monroe, La.			X
Montgomery, Ala.			X
Muncie, Ind.	X	(1)	
Muskegon-Muskegon Heights, Mich.	X	(1)	
Nashville, Tenn.	X	(1)	
New Bedford, Mass.	X	X	

See footnotes at end of list.

SMSA	Mail census ACG	Mail census ACG/DIME	Nonmail census ACG/DIME
New Britain, Conn.	X	X	
New Haven, Conn.	X	X	
New London-Groton-Norwich, Conn.	X	(1)	
New Orleans, La.	X	X	
New York, N.Y.	X	(5)	
Newark, N.J.	X	(1)	
Newport News-Hampton, Va.	X	X	
Norfolk-Portsmouth, Va.	X	X	
Norwalk, Conn.	X	(1)	
Odessa, Tex.			X
Ogden, Utah	X	X	
Oklahoma City, Okla.	X	X	
Omaha, Nebr.-Iowa	X	X	
Orlando, Fla.	X	(1)	
Oxnard-Ventura, Calif.	X	X	
Paterson-Clifton-Passaic, N.J.	X	(1)	
Pensacola, Fla.	X	(1)	
Peoria, Ill.	X	X	
Philadelphia, Pa.-N.J.	X	X	
Phoenix, Ariz.	X	(1)	
Pine Bluff, Ark.			X
Pittsburgh, Pa.	X	X	
Pittsfield, Mass.	X	X	
Ponce, P.R.			(4)
Portland, Maine			X
Portland, Oreg.-Wash.	X	X	
Providence-Pawtucket-Warwick, R.I.-Mass.	X	X	
Provo-Orem, Utah			X
Pueblo, Colo.			X
Racine, Wis.	X	X	
Raleigh, N.C.	X	X	
Reading, Pa.	X	X	
Reno, Nev.			X
Richmond, Va.	X	X	
Roanoke, Va.			X
Rochester, N.Y.	X	X	
Rockford, Ill.	X	X	
Sacramento, Calif.	X	X	
Saginaw, Mich.	X	X	
Salem, Oreg.			X
Salinas-Monterey, Calif.			X
St. Joseph, Mo.			X
St. Louis, Mo.-Ill.	X	X	
Salt Lake City, Utah	X	X	
San Angelo, Tex.			X
San Antonio, Tex.	X	X	
San Bernardino-Riverside-Ontario, Calif.	X	(6)	
San Diego, Calif.	X	X	
San Francisco-Oakland, Calif.	X	X	
San Jose, Calif.	X	X	
San Juan, P.R.			(4)
Santa Barbara, Calif.			X
Savannah, Ga.[3]	X	X	
Scranton, Pa.	X	X	
Seattle-Everett, Wash.	X	X	

See footnotes at end of list.

SMSA	Mail census ACG	Mail census ACG/DIME	Nonmail census ACG/DIME
Sherman-Denison, Tex.			X
Shreveport, La.			X
Sioux City, Iowa-Nebr.			X
Sioux Falls, S. Dak.			X
South Bend, Ind.	X	X	
Spokane, Wash.			X
Springfield, Ill.	X	(1)	
Springfield, Mo.			X
Springfield, Ohio	X	X	
Springfield-Chicopee-Holyoke, Mass.-Conn.	X	X	
Stamford, Conn.	X	(1)	
Steubenville-Weirton, Ohio-W. Va.	X	X	
Stockton, Calif.	X	X	
Syracuse, N.Y.	X	X	
Tacoma, Wash.	X	X	
Tallahassee, Fla.	X	X	
Tampa-St. Petersburg, Fla.	X	X	
Terre Haute, Ind.	X	X	
Texarkana, Tex.-Ark.			X
Toledo, Ohio-Mich.	X	X	
Topeka, Kans.	X	X	
Trenton, N.J.	X	X	
Tucson, Ariz.			X
Tulsa, Okla.	X	X	
Tuscaloosa, Ala.			X
Tyler, Tex.			X
Utica-Rome, N.Y.	X	X	
Vallejo-Napa, Calif.	X	X	
Vineland-Millville-Bridgeton, N.J.	X	(1)	
Waco, Tex.			X
Washington, D.C.-Md.-Va.	X	X	
Waterbury, Conn.	X	(1)	
Waterloo, Iowa			X
West Palm Beach, Fla.	X	X	
Wheeling, W. Va.-Ohio	X	X	
Wichita, Kans.	X	X	
Wichita Falls, Tex.			X
Wilkes-Barre-Hazleton, Pa.	X	X	
Wilmington, Del.-N.J.-Md.	X	X	
Wilmington, N.C.			(2)
Worcester, Mass.	X	X	
York, Pa.	X	X	
Youngstown-Warren, Ohio	X	X	

[1] In ACG program, declined participation in ACG/DIME program.

[2] Declined participation in nonmail program. Census tract coding guide to be prepared.

[3] In mail census ACG and ACG/DIME program although a nonmail census area.

[4] The three Puerto Rican SMSA's are scheduled eventually to be in the nonmail program. However, at this time only the Santurce and Old San Juan portions of the San Juan SMSA are in the program.

[5] Nassau and Suffolk counties are in the ACG/DIME program, remainder of SMSA is not.

[6] San Bernardino County portion is in the ACG/DIME program, Riverside County portion is not in the program.

Metropolitan Mapping Series

Traditionally the Census Bureau has relied on maps supplied by local governments to conduct censuses. Letters were sent to county and municipal governments a year or two prior to a census, asking for street maps that could be used as enumeration maps. Municipal boundaries, enumeration district boundaries, and other census administrative and statistical area boundaries are added to the maps by the Bureau. They were then reproduced for enumeration and supervisory purposes for the census. This procedure, although cumbersome, did allow the census to be taken more or less efficiently, at least in the less densely populated areas of the Nation. For the 1970 census, this procedure will be used in those portions of the country outside the urbanized areas covered by the Metropolitan Mapping Series (MMS) program.

However, the Census Bureau found while conducting the 1960 census that it was becoming difficult to administer the census because of the great number and differing scales of maps received from the communities in these urban centers. Problems of overlapping maps and control of enumeration districts were endemic. As soon as the census was concluded, proposals were developed to create series of street maps at a standard scale for use in censuses and surveys. These proposals developed into the metropolitan mapping series program.

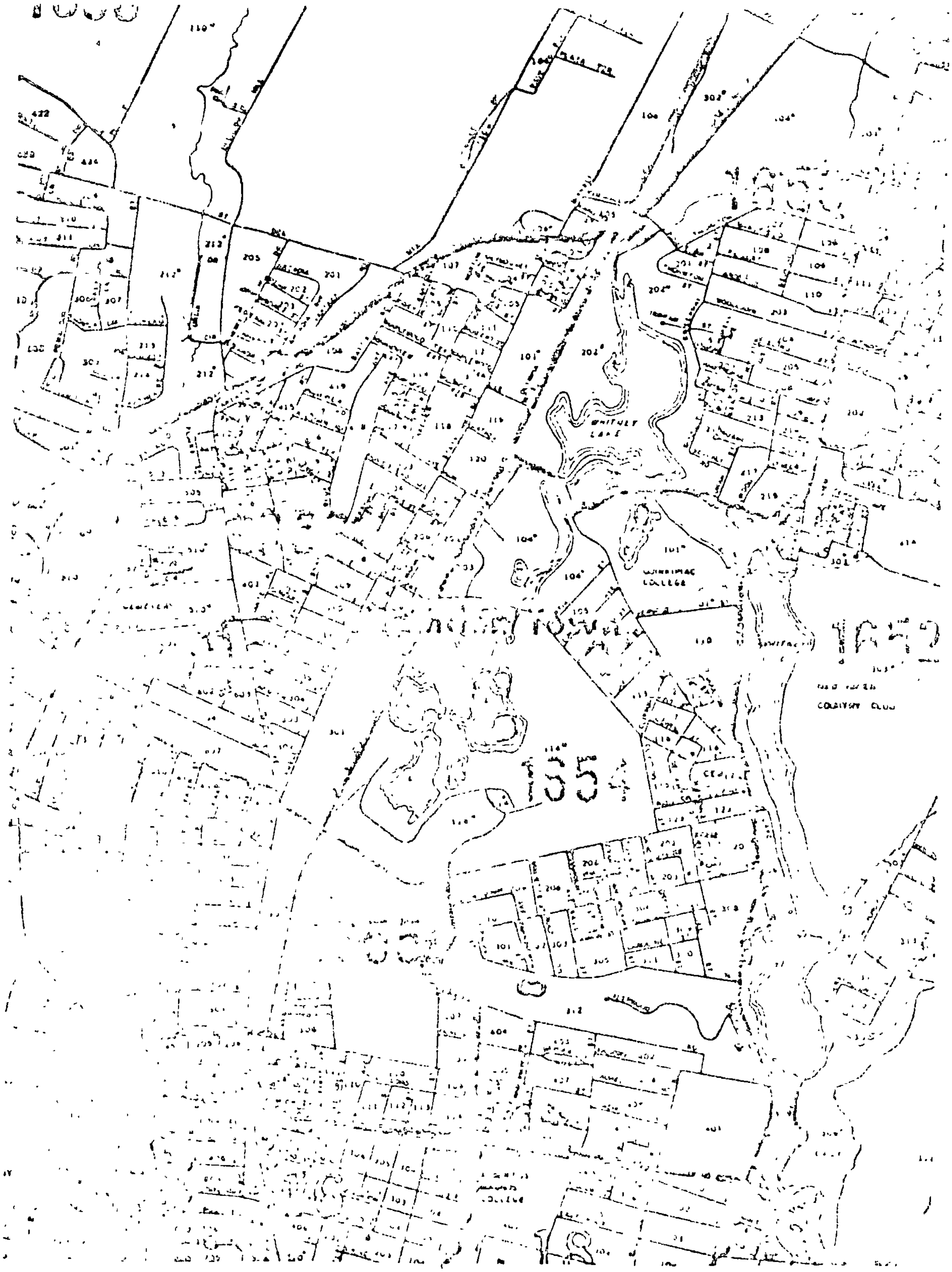
For this program two basic inputs were needed: U.S. Geological Survey topographic maps for control purposes and maps provided by local planning agencies or municipal governments for defining current street patterns, street names, and other pertinent features.

The early maps produced in the MMS program were created from these two sources by Bureau cartographers at a scale of 1 inch = 880 feet. This was later changed to 1 inch = 800 feet. The maps are intended for statistical purposes and therefore are not of engineering quality, although they are generally sufficiently reliable for most planning and administrative purposes. The maps were designed to contain all existing vehicular thoroughfares and their names, railroad tracks, major drainage features, shorelines, lakes, and prominent landmarks that occupy large areas, and have clearly defined boundaries. They also contain State, county, congressional district, minor civil division or census county division, municipal, and census tract boundaries. Block numbers, census tract numbers, State and county names, and place names and codes are also contained on the maps. Maps have presently been drafted for the urbanized portions of all 233 SMSA's. Figure 39 illustrates a portion of an MMS map at half scale (i.e., 1" = 1600').

Once the maps were developed it became imperative to develop an updating procedure so that the maps would be reasonably accurate at census time. The decision was made to involve county and regional planning agencies in this program because of their obvious interest and expertise in the field and their metropolitan-wide coverage. The Department of Housing and Urban Development recognized the usefulness of these maps to planners and made available "701" grant funds for local review. The appropriate agencies were contacted, and all cooperated in the local review process.

The MMS maps will be used for administrative purposes in the 1970 census and basic reference sources for the local coding phase of both the ACG and ACG/DIME geographic base file programs. A reduced scale version (1" = 2000') of the maps will be published for use with the 1970 census results. Preliminary maps at both scales are available at a nominal cost. Final maps will begin to become available in late 1970 or early 1971. A catalog listing available Census Bureau maps, including MMS maps, will be issued during 1970. Inquiries concerning the MMS and other Bureau mapping programs should be directed to Central Users Service, Bureau of the Census, Washington, D.C. 20233.

The MMS will be updated after the 1970 census for future censuses and surveys. Coverage of the series may be extended so that they will be of greater use, not only to the Census Bureau but also to map users around the country.



Fi... ...etropolitan m..pping series

Data Available From the 1970 Census

A variety of data products and services will be available from the 1970 Census of Population and Housing data base. A set of standard tabulations will be available in a variety of media including microfilm, punched cards, printed reports, and computer tapes. Public use samples containing data for samples of individuals from the basic record tapes, but without identifying information, will also be avail- . In addition, other services may be obtained at user request and expense, including special tabulations of the basic record tapes, and computer-generated analytical reports.

As in earlier censuses, printed reports will be the end product of the 1970 census and will be available for purchase and in libraries. Printed reports offer the two main benefits of low cost and easy accessibility. On the other hand, printed reports will not be available until several months after the tabulations have been completed; and they will contain restricted ographic and subject matter detail.

A major element of the 1970 data delivery system will provide magnetic tapes of summary census data. These will have essentially the same subject matter scope as the printed reports, but with greater detail. The summary tapes will begin to become available in the summer of 1970.

The first summary tapes available will be the First Count Summary Tapes, containing final complete-count population and housing data for States, counties, congressional districts, minor civil divisions, places, and enumeration districts or block groups. The subjects tabulated will include age, sex, color, marital status, relationship to head of household, population under 18 and over 65 years of age by household relationship, family type, crowding, tenure of occupied housing units, vacancy status, units in structure, rooms, plumbing facilities, basement, telephone, and value or rent. In addition, many tabulations are cross-classified by race.

Second Count Summary Tapes, containing complete-count population and housing data at the level of census tracts, minor civil divisions, and larger census areas, are expected to begin becoming available in late 1970. These tapes will include the same subjects as the first count tapes, but in much greater detail.

Third Count (Block) Summary Tapes, containing complete-count population and housing data for city blocks, are expected to begin becoming available in the spring of 1971. These tapes will contain fewer data than the tapes for the first or second counts; however, substantially more data will be carried on the block tapes than was available in 1960 or is expected to be printed in 1970.

Fourth Count Summary Tapes, including data down to the census tract or minor civil division level, are expected to be available in mid to late 1971. They will contain 20-, 15-, and 5-percent sample population and housing characteristics such as occupation, income, education, and household equipment and facilities.

Fifth Count Summary Tapes, which contain population and housing sample data summaries for ZIP code areas, are expected to be available in late 1971 or early 1972.

Since summary tapes must be created by the Bureau of the Census in order to process the census, the tapes for all five counts will be available to users at the cost of about $60 per reel. Sets of summary tapes may be ordered for each count by State. The number of reels varies by State and by count. The Third Count for California, for example, will be about 30 reels in all.

Microfilm will also be a generally available medium for obtaining general census tabulations. Printed reports issued from the 1970 census will be available on microfilm. The contents of the summary tape files will also be available in this manner.

An extended array of sample data on magnetic tape for public use is planned for the 1970 census. These tapes will be made available to users at the cost of reproduction, together with appropriate documentation. Although public use samples are limited in geographic detail, they are most useful to researchers interested in the relationships of demographic variables for individuals rather than aggregates.

Special tabulations of census basic record tapes will also be made available. For this service, computer programs are created to meet the specific needs of the user in obtaining data summaries for geographic areas not recognized in general tabulations and/or including subject breakdowns or cross-classifications not appearing in general tabulations. Users are charged for planning, programing, clerical, machine, or other costs necessary to provide this service.

Another type of special service is matching studies. Matching studies take a series of individual records furnished by a user, link them to their respective census questionnaires, and prepare summaries of the census characteristics of the individuals. Individual data are never released.

Other special services for the 1970 census may include computer-produced analytic reports, computer graphics, and software packages for census data analysis. Further information on 1970 census data products and services can be obtained from the Central Users Service, Bureau of the Census, Washington D.C. 20233.

BUREAU OF THE CENSUS

George Hay Brown, Director

Robert F. Drury, Deputy Director

Joseph F. Daly, Associate Director for Research and Development

Robert B. Voight, Special Assistant

The Census Use Study is under the direction of Caby C. Smith, Project Director.

This report was prepared by George Farnsworth and James P. Curry of the Census Use Study. Documentation and the principal research activities related to the development of the DIME system were the responsibility of James P. Corbett, Statistical Research Division, Bureau of the Census; and Donald F. Cooke, George L. Farnsworth, William H. Maxfield, and R. Ross Hall of the Census Use Study Staff.

The author wishes to acknowledge the helpful comments and assistance of V. V. Almendinger, System Development Corporation; Robert E. Barraclough, Department of Transportation; Prof. William L. Garrison, University of Pittsburgh; and Edward F. R. Hearle, Booz, Allen and Hamilton, Inc., each of whom reviewed earlier drafts of this report.

Mr. Hearle served as general editor and coordinator of the report series, assisted by Mr. Curry and Michael G. Garland of the Census Use Study staff. Editorial assistance and report planning were provided by Maureen Padgett, Administrative and Publications Services Division, Bureau of the Census.

SUGGESTED CITATION

U.S. Bureau of the Census, *Census Use Study: The DIME Geocoding System*, Report No. 4
Washington, D.C., 1970

For sale by the Bureau of the Census, Washington, D.C. 20233
or any Department of Commerce field office.
Price 50 cents.

UNITED STATES DEPARTMENT OF COMMERCE PUBLICATION

C2 # census use study

report No 4, July 1970

K 452, Geographic Information System

A. U S - Census

report no. 4
The DIME Geocoding System

T

C1 U.S. DEPARTMENT OF COMMERCE
Bureau of the Census

The Census Use Study, a small-area data research study sponsored by the Bureau of the Census, was established in New Haven, Conn., in September 1966. It was established to explore the current uses and future needs of small area data and data handling and display techniques in local, State, and Federal agencies.

The study was charged with the following objectives:

The development of a system that would allow efficient interrelating of Census Bureau data with other local and State data to meet specific needs.

To investigate the benefits of cooperative data collection between the Census Bureau and other local, State, and Federal agencies.

To investigate the level of detail and the form in which census data should be made available to local users.

To develop computer programs for use by local communities to allow rapid conversion of census data into information useful for local analysis.

To analyze the results of the study for potential procedures to be incorporated in local community programs to take advantage of census and other information.

To publicize the results so that other areas around the country may benefit from the efforts of the study.

A special census of New Haven, conducted in April 1967 by the Bureau of the Census to test proposed 1970 census procedures, provided a basic source of data. Local agencies also made available certain data from their own records to enable testing of data handling techniques developed by the study.

In response to the established goals of the study, exhaustive research was carried out in the following areas:

Geographic base systems
Record matching
Computer mapping
Special tabulations of data
Special sample surveys of family health and area travel patterns
Local data user interests and needs

The study was supported financially by the following Federal agencies: *Department of Commerce; Office of Civil Defense of the Department of the Army; Department of Health, Education, and Welfare; Department of Housing and Urban Development; and the Department of Transportation.*

The city of New Haven provided substantial facility and personnel support, and 30 local agencies cooperated in the study.

The results of the study, including documentation of the computer programs and other procedural guidance, are presented in a series of reports and computer packages listed inside the back cover of this report.

In July 1969, the Southern California Regional Information Study (SCRIS) was established in Los Angeles, Calif. The study is jointly sponsored by the Bureau of the Census and the Southern California Association of Governments, an association of local and county governments in the Los Angeles area. SCRIS, of which the Census Use Study forms the Bureau's contingent, will attempt to transfer experience gained in New Haven to a larger urban area, with a view to assisting census data users in all large urban areas in preparing for and using 1970 census data as it becomes available.

Reports

1. *General Description.* An overview of the development and operations of the New Haven Census Use Study.
2. *Computer Mapping.* A report on the mapping of census and local data using several computer mapping techniques.
3. *Data Tabulation Activities.* A report on the contents and uses of special tabulations provided to local agencies from the 1967 special census of New Haven, Conn.
4. *The DIME Geocoding System.* A report on the development of the DIME geographic base file including a description of the file and the unit system, uses of the file, and methods for creating a DIME file.
5. *Data Interests of Local Agencies.* A description of a series of surveys undertaken to explore the needs of local agencies for small area data.
6. *Family Health Survey.* A report on a sample survey taken to augment data from the special census of New Haven with information on various elements of family health.
7. *Health Information System.* This report documents the development of a maternal and child health information system utilizing census and local data.
8. *Data Uses in Health Planning.* This report outlines the uses of data in health planning based on the general research conducted at the Census Use Study.
9. *Data Uses in Urban Planning.* A description of the general findings of the Census Use Study as they apply to the field of urban planning.
10. *Data Uses in School Administration.* A report describing the uses of data in school administration based upon activities conducted at the Census Use Study with local school administrators.
11. *Area Travel Survey.* A description of a sample survey conducted to augment the New Haven special census data with basic data for use in transportation planning.
12. *Health Information System - II.* A supplementary report on the development and implementation of a computer-based health information system including introductory materials, documentation, methodology, and analysis.
13. *Community Shelter Planning Project.* A report describing the development of a computerized system for allocating people to fallout shelters using census data and Census Use Study tools.

Computer Program Packages

ADMATCH: An Address Matching System. A computer program package designed for use in assigning geographic codes to local records using a DIME or similar geographic base file. Includes a users manual and computer programs.

DIME: A Geographic Base File System. A computer program package for creating a DIME geographic base file. Includes clerical instruction, a computer manual, and programs.

GRIDS: A Computer Mapping System. A computer program package for use on small-scale computers which provides three mapping options within a grid pattern: density, shading, and value maps. Includes users manual and computer programs.

All reports and programs described above are scheduled for release in 1970. For information on specific titles, write to Publications Distribution Section, Bureau of the Census, Washington, D.C. 20233.

oratory for Computer Graphics and Spatial Analysis
Graduate School of Design Harvard University

Allan H. Schmidt
Acting Director
Eric Teicholz
Assoc. Director

INTERACTIVE MAPPING OF URBAN DATA

Eric Teicholz
Associate Director.
Laboratory for Computer
Graphics & Spatial Analysis,
Harvard University.
Graduate School of Design

April 1975

This paper originally appeared in the Proceedings of the Second General Assembly of the World Future Society, June 1975.

I. ABSTRACT

The article describes several factors that are contributing to the explosion of statistical and geographical data related to our urban areas and the corresponding interest in automated procedures for the input (capture), analysis and display of spatial data.

Basic automated mapping methods and procedures are described and illustrated using two interactive mapping systems called INPOM and ASPEX, developed at the Laboratory.

Finally some limitations of computer mapping and cost considerations are described.

II. INTRODUCTION

In a recent issue of the New York Times, an article appeared stating that Canada was "going metric" and that this event is expected to increase pressure on the last major non-metric holdout - the United States.

One ramification of this conversion is that large numbers of existing manual maps will become immediately obsolete. Another possible result of this drift towards metric conversion is that the Laboratory for Computer Graphics and Spatial Analysis (the Laboratory) is receiving numerous requests, primarily from utility companies and planning agencies, requesting information on how to go about developing automated techniques for the collection (data capture), analysis and display of spatial data.

There are other reasons why there is a great deal of interest in automating procedures relative to the analysis and display of urban information. There is a large increase in both the volume and quality of statistical data This is due to automated techniques being used by traditional agencies such as the Department of Commerce (who collect and distribute Census data) as well as a result of a host of other governmental and private agencies using computers for collecting statistical data along with geographic identifiers - thereby creating geographic entities which can be used for a variety of analysis and display programs. The Central Intelligence Agency, for example, distributes political boundaries for all countries (World Data Bank I) and will soon release much more detailed breakdowns (World Data Bank II). The United States Geological Survey (USGS) has a multi-million dollar commitment to automate the National Map Series and they claim that all the USGS topological maps will be available in digital form in five to ten years.

Along with the increasing amount of available geographical and statistical data is a demand for new, more detailed and more accurate data on the part of urban researchers. New integrated hardware and software cartographic "turnkey" systems make this data more available and cost beneficial than ever before. Commercial companies such as Computervision (Bedford, Mass.), Applicon (Burlington, Mass.), Calcomp (Anaheim, Cal.), and Calma (Sunnyvale, Cal.) offer such systems. In addition, most major time sharing companies now support remote graphic applications encouraging the development and use of interactive mapping. Finally, dissemination groups such as the Laboratory, the Census Bureau and the Geography Program Exchange (East Lansing, Michigan) are distributing low cost graphic display programs.

III. BACKGROUND

There are several classes (types) of maps that are used for the mapping of spatial data. Base maps display geographic entities such as boundaries (Census tracts, Standard Metropolitan Statistical Areas, blocks), road and river networks or almost any of the twenty-odd cartographic features that are overlayed to produce a USGS topological map. Base maps are normally used to convey locational data but do not convey other types of statistical information.

Thematic maps display geographical concepts such as gradients, density distributions, magnitudes of various attributes or other quantitative or qualitative data. To display geographical concepts, a variety of techniques are employed such as various types of symbolisms, grey tones and color symbols and tones - all of which can be superimposed on a base map.

Statistical surfaces can be represented using chloropleth or isarithmic maps. The former represent statistical variables by conforming to a particular boundary or enumeration district. Input will consist of polygon coordinate data and statistics that relate to the geographical areas. Isarithmic mapping emphasizes gradients such as contours or other isarithms to represent areas and volumes to portray a continuous real (or assumed) statistical surface.

An additional type of map that should be mentioned is called a cartogram which deliberately distorts areas or volumes to represent an aerial quality. The example below, taken from an article by R. L. Phillips in the April 1974 Proceedings of the IEEE (Vol. 62, No. 4, p.442) illustrates a retail market view of the United States employing a program that produces cartograms.

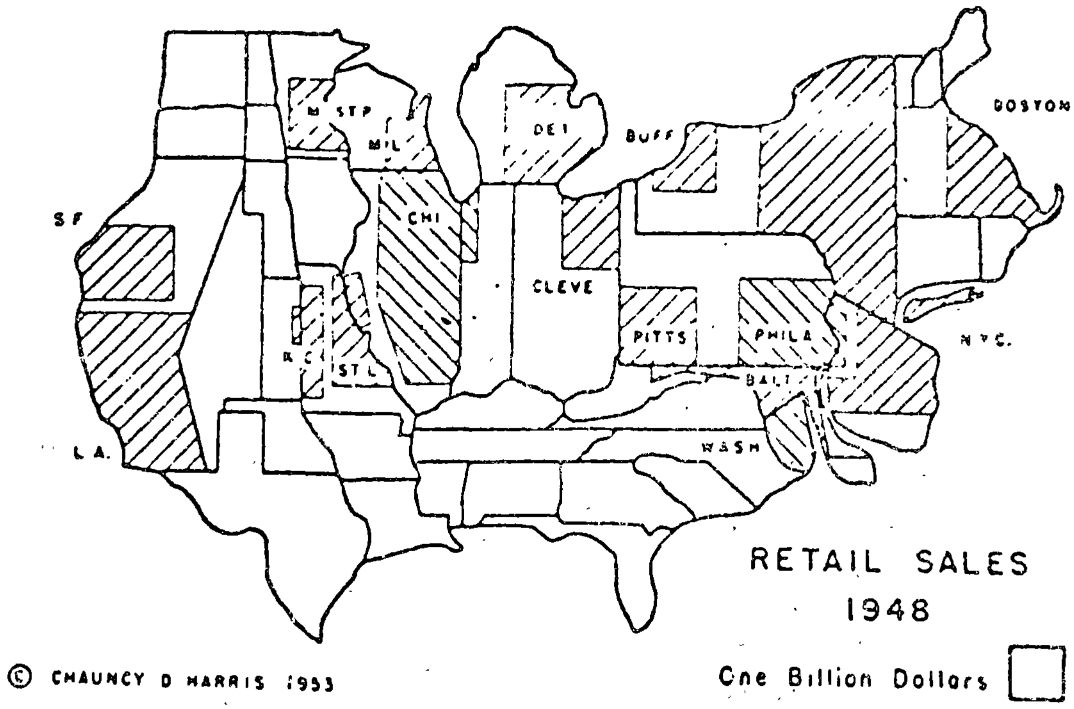


Figure 1: Cartogram Illustrating Retail Sales in the U.S. for 1948

IV. INTERACTIVE MAPPING: The ASPEX and INPOM Programs

Perhaps interactive computer mapping offers one of the most powerful tools for urban researchers to date. One can perform on-line data editing operations using an intermediate display device such as a cathode ray tube (CRT) (either color or black and white), and can selectively retrieve and "massage" data to perform a variety of statistical operations. A user can then alter values, class intervals, symbol and shading types of the output display map. Finally, a user can transform the data, look at each view on a CRT until the desired result is achieved, and then output the final display file to a variety of display hardware devices such as digital plotters, COM (computer on microfilm) plotters, a color matrix plotter (such as the new Color Jet Plotter from Sweden), photoplotters or some other output device. In some cases the resolution of the cathode ray tube itself might be satisfactory. The ASPEX and INPOM illustrations appearing in this article were reproduced from a Tektronix 4610 hard copy output.

At the Laboratory, a variety of ongoing research projects are involved in the interactive capture, processing and display of spatial data.

INPOM

The Interactive Polygon Mapping System (INPOM) is designed to produce maps of countries, states, census tracts and other arbitrarily shaped regions defined within a geographic base file (GBF). INPOM is a two-dimensional mapping program capable of producing conformant base and thematic (shaded) maps. It has the capability of selectively retrieving areas to be mapped, of controlling the degree of detail to be displayed (for outlines) and in the symbolism used to depict data values. The user can zoom in on particular areas of interest, try different types of symbolism, get immediate hard copy from the CRT display, and vary the amount of detail to be displayed.

The flexibility of input is achieved by entering keyword-type commands from the display terminal. The program responds by requesting additional information needed to execute the command. Because of the internal data structure used by the program, it is possible to get listings of points coordinates, chains (the data structure used by the program), single polygons, or user defined regions within the study area.

At present, there are over 30 input commands operating in conversational mode. The commands are entered as 2, 3, or 4-letter nmemonics and the program will respond accordingly by requesting numeric data or alphabetic responses. The numeric responses may be values, coordinates (which may be stored internally if desired), window parameters and the like. All data is free field format so that the user does not have to worry about restrictive fixed field formatting requirements.

Another flexibility of INPOM is that all commands have default conventions or values which the program will preset for the parameters of a command until the user employs that command. Once set, the parameters of a command remain in effect until the command is again specified. Figure 2 below shows the current command file for the INPOM program. Figure 3 (detail level = 1) is a base map of Africa while Figure 4 (detail level = 5) shows a thematic map of Africa showing the gross national product on a per capita basis based on 1970 data. Figure 5 (detail level = 10) shows the same data illustrating the zooming and increased detail level for a section of West Africa.

File Input, Creation and Querying

INC	(INput Chains)
INP	(INput Polygons)
VAL	(Input VALues)
FN	(File Name)
RV	(Read Values)
WV	(Write Values)
LIN	(define LINe legends)
CVRT	(input Chains from the POLYVRT program)
REN	(REName chains and or polygons)
INFO	(list INFOrmation on chains, polygons, lines)

Value Level and Symbolism Definition

NL	(Number of value Levels)
LVL	(define value LeVeLs)
SHD	(define SHaDing for levels)
FAS	(FActor Shading density)

Termination

EXIT	(EXIT from INPOM to monitor level)

Windowing, Scaling and Polygon Selection

WORG	(Window ORIGin location)
WSIZ	(WIndow SIZe specification)
MW	(Move Window [illegible])
PMM	(Polygon selection by Minimum and Maximum extents)
DMM	(Data coordinate Min-Max selection of polygons)
FDW	(Fill Data Window with partial chains)
FSW	(Fill entire Screen Window with polygons and chains)
FA	(FActor for expanding or shrinking map)
XFM	(Transform point coordinates with respect to a location)

Graphic Manipulation Instructions

OL	(to specify OutLine or shaded maps as output)
DET	(highest DETail level to be drawn)
DL	(Draw Line legends)
DWO	(Draw Window Outline)
PLT	(PLoTing mode (for Tektronix 4014 display only))
MAP	(draw a MAP)

FIGURE 2
Command File for INPOM

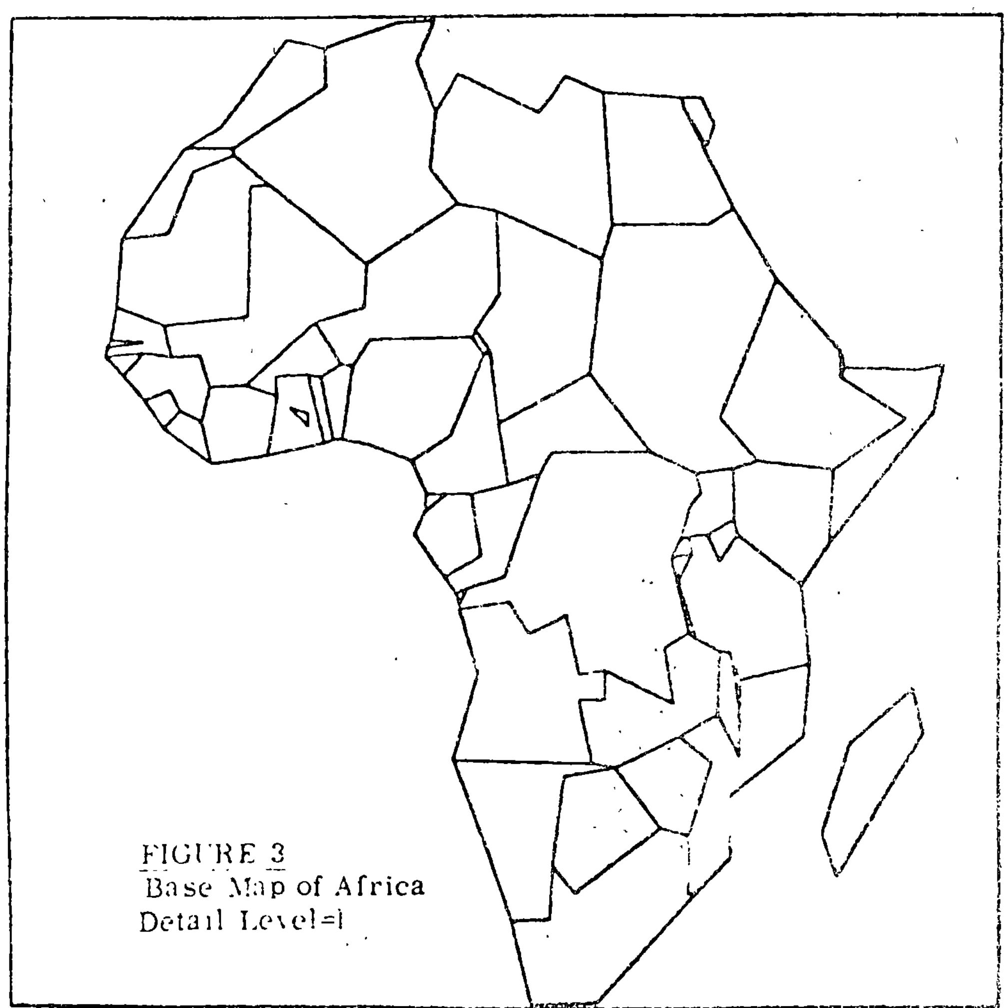


FIGURE 3
Base Map of Africa
Detail Level=1

WEST AFRICA

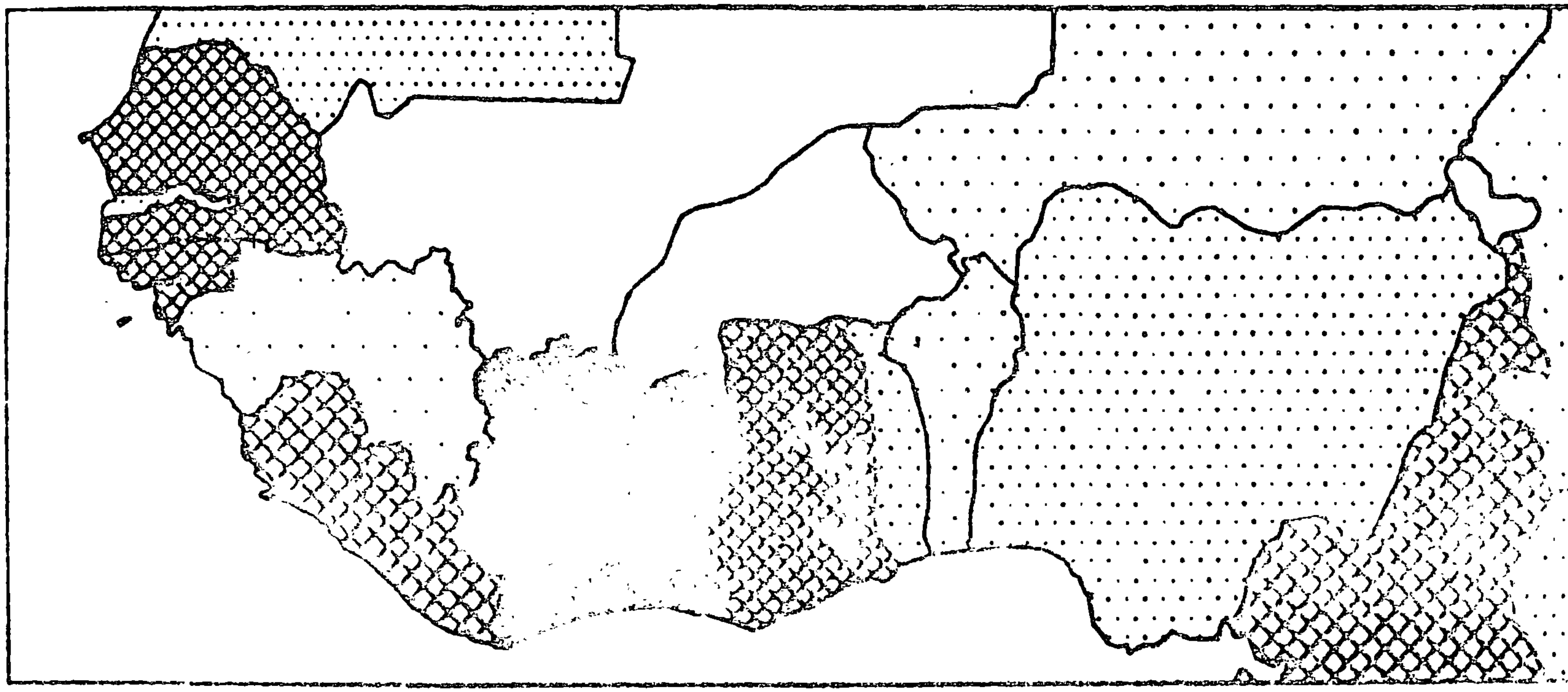


Figure 5

Detail of West Africa
Detail Level = 10

BY INPOM

ASPEX

The Automated Surface Perspective Program (ASPEX) is an interactive (a batch version is also being developed) program that displays three-dimensional representations of statistical surfaces. Such representations of three variables were not very common until the advent of the computer because of their difficulty in construction. Although most people are not particularly accustomed to reading information displayed on surfaces (especially when the information is statistical or mathematical in nature), mathematicians, cartographers and planners are beginning to accept three-dimensional surface representation as a powerful extension of two-dimensional mapping.

The ASPEX program takes a matrix (or array) of data of any size. The program incorporates a free field format command language that operates on nmemonic keywords for the over 70 commands of the program. The commands deal with the following categories:

- -initialization and production (such as display, help, define, expunge, plot, etc.)
- -data input and storage (number of columns, data type, grid input, header information, etc.)
- -data value manipulation (min, max, smoothing, square root, etc.)
- -viewing parameters (including view type such as isometric, planometric, and perspective) and orientation
- -graphic options (including data surface commands such as draw, height, interval, symbol size and cosmetic features such as base information, map scale, title, etc.)

Another important flexibility of ASPEX is the ability to alter the viewpoint so that a user can be located anywhere beyond, above, or upon the surface. Capabilities are also being added to draw features directly on the surface. The scaling of the output plot may be to any predetermined height, width or window size and is accomplished automatically by the program. The three figures below represent different views of the U.S. but generated from the same data base.

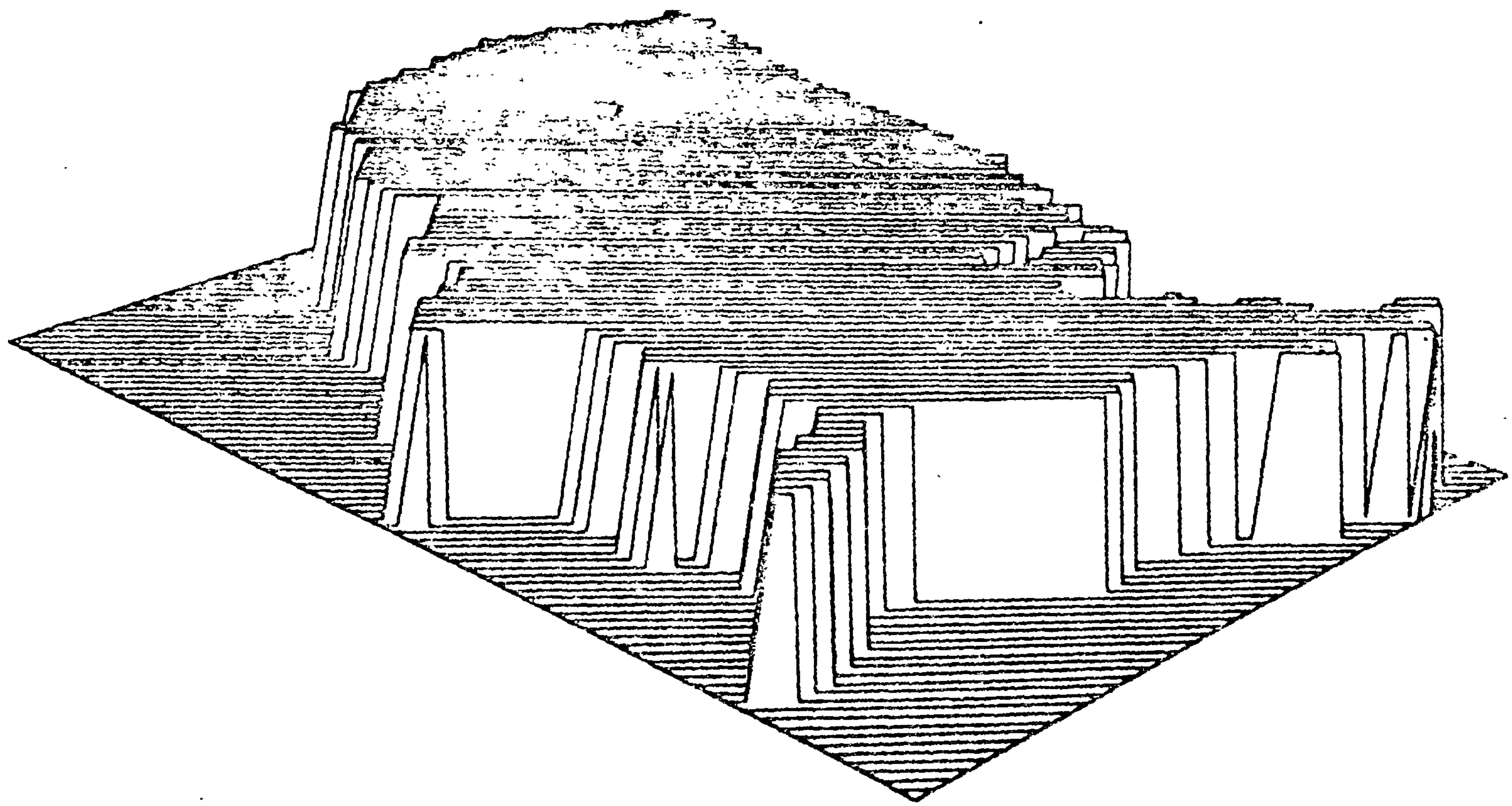


Figure 6

Truncated Height Values

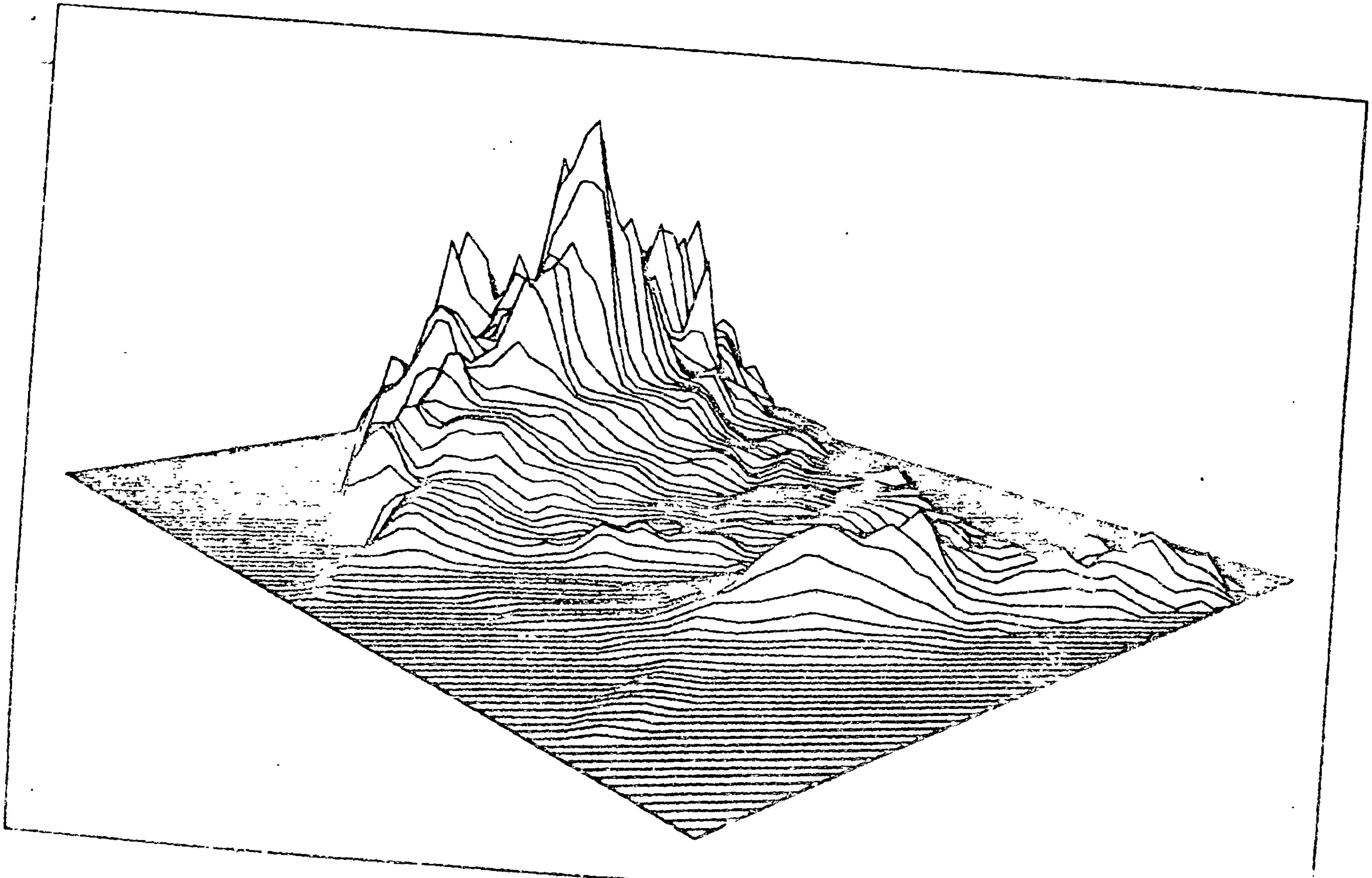


Figure 7

View from the South East

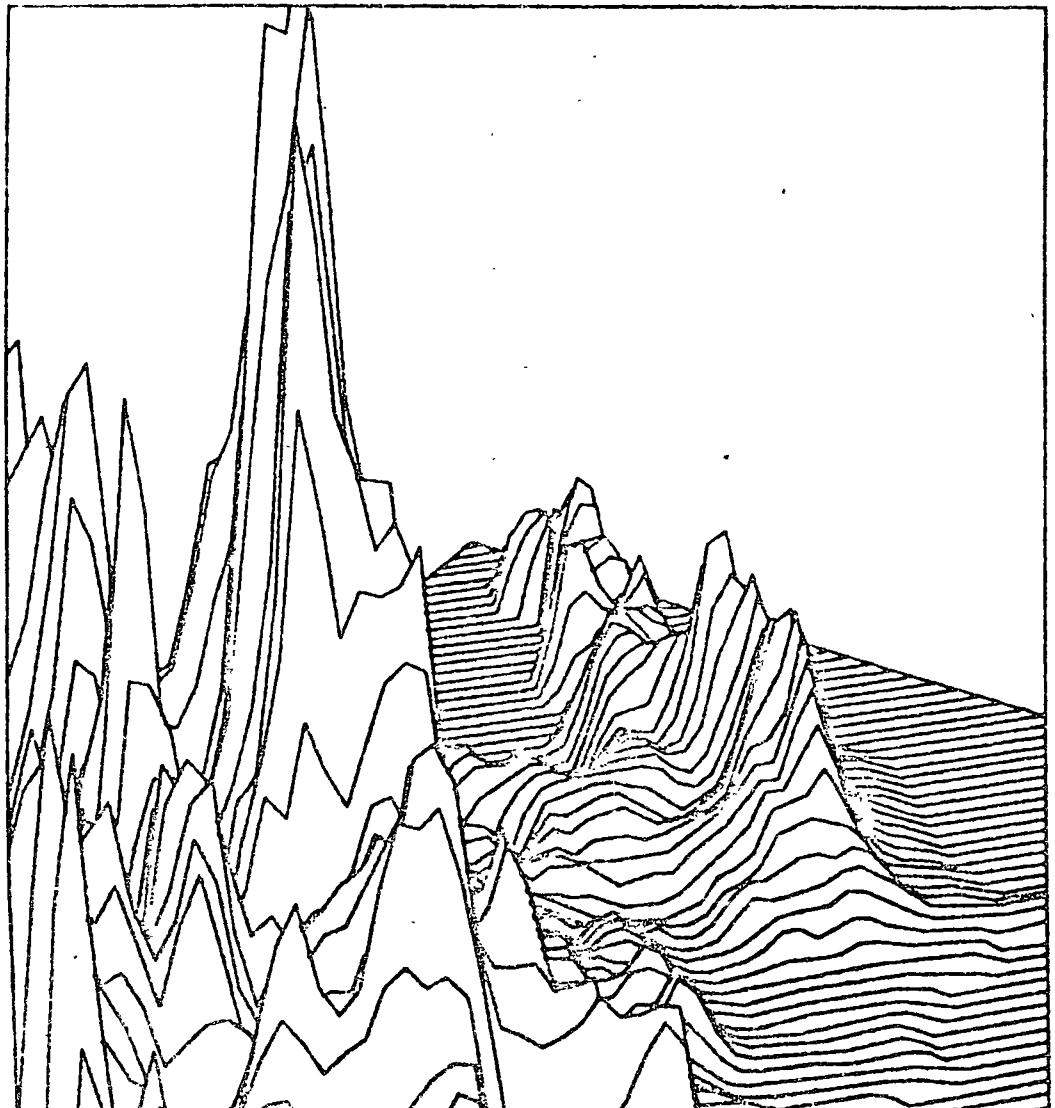


Figure 8

View from the Rockies

V. CONCLUSION

The software to rapidly and economically generate computer maps is clearly available. Unfortunately, this does not necessarily make automated mapping a pragmatic analytic tool. The economics of computer cartography must take into consideration the totality of the process. Questions on input such as where the data is coming from, its reliability, and updating procedures must be considered. One is reminded of the millions of dollars that was spent on data banks in the 1960's that became data dumps in the 1970's.

One must also take into account GBF's from non-contiguous sources that are in different formats. There are also central processor considerations which must reduce the data to its lowest common denominator, restructure it, process it and finally output a display file for a particular output device.

Output devices range from the ubiquitous line printer which produces inexpensive low-resolution maps using over-printing techniques to digital line plotters to COM (computer on microfilm) and expensive photoplotters.

A final word of caution relates to the maps themselves. Being able to produce maps efficiently and inexpensively is no guarantee that the maps will facilitate and improve the decision-making process of urban researchers. One must be able to clearly understand what is being represented by a map and the purpose for which it is intended. Only then will the application of computer technology have any meaning.

COMPUTER APPLICATIONS IN LANDUSE MAPPING AND THE MINNESOTA LAND MANAGEMENT INFORMATION SYSTEM

Mei-Ling Hsu, Kenneth Kozar, George W. Orning, and Pamela G. Streed

The concept of rational use of land resources is not new, but the general public's awareness of the limited nature of our land resources and the urgent need for optimal utilization of them are phenomena of this decade. Today, most planners, officials, and the public share this awareness. In order to manage land resources effectively and plan for the future, it is essential that we begin to comprehend the land utilization of the past and present. To achieve this, we are in need of a vast amount of information on land resources and related socio-economic variables. When one is dealing with a large area such as a major region or country, automation becomes necessary for carrying out the tasks of information collection, analysis, and updating. Accordingly, the State of Minnesota has established an automated land information system to facilitate the work of resource management and planning.

The Minnesota Land Management Information System Study, MLMIS, is being developed under the auspices of the Minnesota State Planning Agency and the University of Minnesota Center for Urban and Regional Affairs. The system is primarily a result of the four earlier studies which began in 1966: (1) a study on lakeshore development of a small area in central Minnesota, the Brainerd area in Crow Wing County (Orning, 1967); (2) an expansion of the Brainerd study, the Minnesota Lake Shore Development Study (Borchert, et al., 1970a and 1970b); (3) a report on state land holdings (Minnesota State Planning Agency, 1968); and (4) the state land use mapping project (MLMIS, 1971). During the undertaking of these studies, it became apparent that a statewide data system was essential to these types of investigations and a rational use of land resources. Therefore, the MLMIS aims at providing extensive information to officials, planners, and researchers in decision-making and policy formulation concerning Minnesota land and water. The study works toward establishing a statewide data base for land related information, i.e., landuse, land ownership, land value, land characteristics, and government landuse controls.

In the following, this paper will discuss the system design and current developments of MLMIS. These include the topics of basic data unit, the geocoding system, major information sources, data structure, procedures of data input, procedure of data retrieval, display and analysis of land information, and finally data updating and potential information source. It should be noted at this point that the MLMIS does not claim superiority in hardware or software design. The system has been funded modestly by the state, and thus far has only employed a handful of personnel trained in systems management. But it is progressing rapidly and, more importantly, it has been producing valuable results such as landuse maps and research reports on land resources and utilization. In addition it has provided data and consultation to a number of state agencies and private firms.

The Basic Data Unit

The choice of a basic data collection unit is crucial to an information system. In Minnesota, it was determined as a result of much use and examination of land records that the basic unit for a statewide coverage should be the forty-acre parcel, the smallest consistent unit in the U. S. land survey system. This system was employed to survey most lands in the central and western United States. The first order reference axes in the system are the 32 pairs of locally defined principal meridians and base lines (parallels). The second order references are tiers of townships and ranges (Figs. 1 & 2). A township, which is defined by a pair of township lines and ranges, is a 36-mile square. Within a township, there are 36 sections (one-mile square), and in each section, 16 forty-acre parcels.

There are nearly 1.4 million of these forty-acre parcels or data-cells in Minnesota. Most blocks of land, whether in public or private ownership, have the edges of forty-acre parcels as boundary lines. The parcel divisions are reflected in agricultural areas as field lines, in forested areas as timber cutting boundaries, and in cities as major streets. In fact, in cities, major commercial developments often occur on section corners of the U. S. land survey. These lines and cells describe the manner in which people have divided and shaped the landscape of Minnesota. The pattern is quite evident when the state is viewed from the air (Fig. 3). The forty-acre parcel is important in other ways. This parcel is recorded on many governmental records for Minnesota at the county, regional, state, and federal levels. The cell structure also lends itself to computer mapping which is necessary in analyzing large volumes of data over extensive areas.

It may be argued that the forty-acre parcel is not a good choice for the basic data unit. The U. S. land survey system is based on locally defined reference axes, and therefore does not have global applications. With this respect, a better choice would be the Universal Transverse Mercator System (UTM). The UTM is well defined on large scale topographic maps. From the geographic point of view, however, the forty-acre parcel is an excellent choice because it is closely associated with the land utilization and settlement patterns of Minnesota. However, the UTM system has been used to record the southwest corner of every township in the state.

The MLMIS is designed to serve mainly planning and studies at the state, regional, and county levels. It is not intended to be a municipal or urban information system, and it is not effective in small-scale investigations, such as those needing information on sewer lines and city streets. Special care, however, is being taken to assure that data collected below the level of forty-acre units can be aggregated and incorporated into MLMIS. Generally, these data are maintained for municipal or county subdivisions. One such example is an urban study employing 10-acre data cells (Robinette, 1971).

The Geocoding System

The locational content of data in Minnesota consists of the following hierarchies: the state, county, township, section, and parcel. The geocoding system in MLMIS contains the identification codes for all these elements. In addition, it encodes the minor civil divisions (MCD) which are the administrative subdivisions comprised of townships, incorporated places, and other areas. The MCD and the county identification are adopted from the U. S. Bureau of the Census; therefore, the MLMIS is linked to the coding system of the U. S. censuses. The centroids of the MCD are recorded in latitudes and longitudes, correct to the nearest 10 seconds. The ability to locate a forty-acre parcel by a point reference to spherical coordinates, however, has yet to be perfected. In the future, one point of each parcel may be identified in latitude and longitude. At present, each forty-acre parcel is uniquely defined by a serial number of fourteen characters, and parcel data may be displayed on township maps to be discussed later (Fig. 2 and Table 1).

In the U. S. land survey system, some townships are not precisely thirty-six square miles. In these irregularly shaped townships, some sections contain more than sixteen parcels per section, and these "extra" parcels are identified by a special code. Approximately ten percent of the townships needed some adjustments.

Major Information Sources

There are four major data sources for MLMIS: (1) various levels of governments, county, state, and federal; (2) the U. S. Bureau of the Census; (3) the University of Minnesota excluding MLMIS; and (4) the MLMIS itself. The Bureau of the Census is listed independently because of its obvious importance. The census data may be used directly by the MLMIS, since the system has adopted the census areal codes, as indicated previously. In contrast, data and records provided by other governmental agencies may be used only after some reorganization of data elements.

Nevertheless, a major purpose of the MLMIS is to make optimal use of land information concerning the State of Minnesota. By building an automatic geocoding and transferring system, the MLMIS attempts to systematize a broad range of data now collected routinely and maintained separately by various government agencies in their licensing and managing functions. Currently, these data can be used only by their collectors, and are not readily available for general use. Most of these data have not been collected or encoded in any compatible manner. In the future, MLMIS may also develop standardized methods of data collection to improve the quality and to facilitate data input to MLMIS.

Various departments of the University of Minnesota have contributed information and their research findings to MLMIS. For example, the Department of Soils has provided information on soil types. Finally, the MLMIS through the Center for Urban and Regional Affairs has collected the bulk of the data now contained in the system.

At present, the main body of data content of the MLMIS is the complete coverage (of 1.4 million cells) of state landuse types and water orientation. In the United States, other data systems comparable to the MLMIS are described by Swanson and Denenberg (Swanson, 1969, Denenberg, et al.). However, neither has this large a number of data cells. Information on water orientation was obtained from aerial photographs and the county highway maps issued by the state Highway Department. It contains seven types of orientations denoting wherever a forty-acre parcel adjoined a lake, stream, or ditch. Information on landuse was interpreted from aerial photographs. A nine-category classification was established so that the interpreters could determine the landuse types from the photographs with a minimum consultation of other data sources (Orning and Maki, 1972, Appendix C). The nine categories are: forested, water, marsh, cultivated, pasture and open, urban residential, urban non-residential or mixed residential, extractive, and lastly, transportation. The dominant landuse was identified for each forty-acre parcel from the photographs which were read stereoscopically by a three-man team, two interpreters and one map recorder. Double interpretation facilitated accuracy. The basic unit of interpretation was a township within which section lines were followed. Each regular section was divided into sixteen forty-acre parcels based on field lines, timber cut edges, as well as a transparent reference grid. It was assumed that these parcels cover all surface area, including water bodies. Despite the clearly defined landuse categories there were many cases in which it was difficult to determine from the photographs the dominant landuse. These problems were resolved by field checking.

MLMIS Data Structure

The data are presently maintained on twelve machine readable magnetic tapes; with one exception, each tape contains one development region defined by the State of Minnesota. Development Region 3 is on two tapes. A development region contains one to eleven contiguous counties. All tapes are maintained on the CDC 6600 computer system under the 7000 MODIFY/SCOPE OPL format with 556 BPI. Each of the eighty-seven counties in the state is contained in a separate file on the tape corresponding to its development region. The county field consists of records for each forty-acre parcel or government lot.

Each unique record of a parcel has the following four components: a 14-character identification key, minor civil division (MCD) number, longitude and latitude of the MCD centroid, and seven types of data. These data types are landuse, water orientation, federal land ownership, relative ownership (full or partial of the parcel, federal ownership only), geomorphic region, state land class, and soils. The last three types are completed for only a small number of counties.

A brief description on the hardware situation of the MLMIS is in order, since the hardware affects the designs of data structure and input/output procedures. The MLMIS has been a very economical system with respect to equipment purchasing (Table 2). Except for the last item, these machines are the property of the University Computer Center. The Image Analyzer was funded by several sources.

The Procedures of Data Input

The procedure of data input to MLMIS employed before 1973 was a traditional one. It contains the following steps. For each forty-acre parcel, land information was coded on maps (scale 1:24,000), locational and land information was recorded on mark sense cards, the card information was read, checked, and finally transferred onto tapes. This procedure is cumbersome.

Lately, a new method of data input has been developed called the CRT data entry system, which would eventually improve and expand the MLMIS operations. The CRT (cathode ray tube) is connected to the University's main computer CDC 6600 via a CDC 3200 (Fig. 4). At the CDC 3200, programs and data are stored on disk pack and are available for the work on CRT. At this experimental stage, only one county, Itasca, is programmed into the CRT data entry system. This system consists of two files, locational directory and data files. Data of forty-acre parcels may be called to be shown on CRT for an area one-half of a township at a time (Fig. 5). This areal coverage on screen is limited by the CRT (CDC 211) capacity which allows 50 characters horizontally and 20 vertically. However, it does provide a comfortable view on the CRT screen. Each parcel is designated by a two-digit code, thus spatially it forms a small square, resembling the "map image" of a parcel on a township map. Subroutines are available for displaying either the data as stored on the computer tapes or a map with coded symbols for one-half of a township. Entering new data, correcting, and updating information can be performed easily, and the results of these operations are viewed immediately on the screen. Hard copies of the final records are produced by the line printer through an instruction from the CRT.

The CRT input procedure is most effective with graphic source materials such as maps and photographs, but it is by no means limited to such usage. In many ways it is superior to the previously employed input method. If the source materials are of comparable scales to the screen imagery, they may be used directly for data input. The results of data entry may be checked immediately by viewing the spatial patterns on the original materials and that on the screen. If only one class of data exists within a township, such as forested land, the data entry may be completed by a single request call on CRT. If it has a predominant class, this may be entered first for the entire township. Then other classes may be input in areas where the data fall under these classes. No punch cards or transient tapes are necessary in this system, for input goes directly to the CDC 3200 disk pack. Lastly, the entire operation is much less dreary, and therefore it moves along faster and with fewer chances for error. Currently, the CRT data entry system is being implemented, and for Itasca County, a dozen new types of data such as county zoning and school districts are entered for each forty-acre parcel.

Procedure of Data Retrieval

At present MLMIS does not have a user oriented program for expedient retrieval of data from the system. This remains to be a weak point in the system. Data are stored on tapes by state development regions and by counties. In order to obtain data on forty-acre parcels of several townships in different counties, for example, a special program would have to be written to search for, first the requested counties, and then the townships, etc. A simple request of data would cost approximately $8.20 at the Minnesota computer system, not including the time for program writing and submission. However, the situation will be improved greatly in the near future. The University Computer Center has purchased a data retrieval package, System 2000. In the future, MLMIS will attempt to employ this system for its data retrieval.

Display and Analysis of Land Information

Computer mapping is the primary mode of data display for

MLMIS. Line printer, Calcomp plotter, and CRT have been utilized for cartographic work. These devices are particularly useful for the large amounts of data analyzed and for displaying land information in a spatial context. Needless to say, statistical programs are also available for data analysis and tabulation.

A MINNMAP program produces maps based on data of forty-acre parcels (Fig. 6). The most substantial contribution of the MLMIS to date in data analysis and mapping is the statewide landuse map at 1:500,000. It is a multi-colored map, thus it cannot be reproduced for this publication. Based on its dominant landuse, each forty-acre parcel is classified into one of the nine categories: forested, cultivated, water, marsh, pasture and open, urban residential, urban non-residential or mixed residential, extractive, and lastly, transportation. The map is a product of sequential use of a computer printer and photographic process. By employing the program MINNMAP, the line printer produced a grey-tone township map for each of the primary colors (red, yellow, and blue). The township maps were then pasted into blocks of six townships, and these blocks were aligned and photographically reduced (Fig. 7). A plate was made for each of the colors required for printing the resultant map. This map, which contains 1.4 million data cells, would have been a difficult task if conventional cartographic procedures were used. But the reproduction method just described was also tedious. As a result other mapping methods are being developed.

Recently, experimentation has moved into the application of the Calcomp plotter to data display. With the plotter a large area such as a county, region, or even the entire state can be mapped in a single computer submission. This eliminates the need for pasting printer outputs of small areas (Figs. 8 & 9). On a state county map, each plotted symbol represents one forty-acre cell. Different numerals, symbols, and/or colors are employed in differentiating symbol classes.

Data Updating and Potential Information Source

The CRT data entry system provides a means of data editing and updating, as discussed earlier. Since land utilization changes through time, information sources and methods for data updating are of utmost importance to a data system like the MLMIS. One potential information source which is being investigated is satellite imagery. Via a grant to the State Planning Agency from NASA, the MLMIS is examining the applications of satellite imagery to Minnesota landuse mapping (Brown, et al., 1973). Earth Resources Technology Satellite (ERT-1) provides the imagery in the form of bulk MSS 70 mm positive transparencies, projected for interpretation of individual bands or color combined. Slides are made of these color combined scenes and projected for interpretation at scales ranging from 1:30,000 to 1:250,000. Nine-inch bulk positive and negative transparencies are being analyzed by density level slicing with an Interpretation System VP-8 image analyzer.

Imagery is evaluated as a potential source for supplying up-to-date information on landuse. It is hoped that the resolution will enable the expansion of the classification scheme currently in use. Thus far, research has shown good results with regard to urban and extractive land uses. In urban places with populations of 7,000 or more, for example, areas of different urban functions can be detected, making possible a refinement of the two-class urban landuse types now employed. In mining areas, a great deal of area measurement and extractive feature classification can be carried out with sufficient accuracy. Research on these and other landuse types are still in progress. In addition, preliminary studies on density of artificial surfaces in the Twin Cities indicate that with some ground truth it will be possible to collect data on impermeability by one-mile cells, and to map the degree of impermeability in the urban area. Moreover, a model for urban run off may be developed based on the data on impermeability and storm sewer networks (Brown, et al., 1973).

MLMIS is an ongoing study; thus far most efforts have been placed in areas of data collection, improvements of input and output procedures, mapping programs, etc. Some progress, however, has been made in data analysis, model formulation, and prediction. Indeed, much of the value of MLMIS has been in the accomplishment of research which meets current needs in land management and planning. A list of major publications is included in this paper. More recent studies include land for development in northern Minnesota (Rusch, et al., 1972), recreational resource study on lakes in the St. Paul area (Wietecki and Orning, 1973), legal controls in relation to landuse (Gilbert, 1973), and the Rapid Analysis Fiscal Tool or RAFT (MLMIS and CURA, 1972). The potential impact of RAFT should be noted. It is a group of computer programs designed to create and manipulate a data base which is to serve as a means of analyzing current laws and proposed alternative policies on state taxation. When it is completed, it will contain a data base relevant to the formulation of fiscal laws and policies, and will be equipped with a package of fiscal models capable of evaluating present and proposed changes to tax laws.

In conclusion, the creation of the MLMIS has been made possible by the promotion of long-term cooperation and coordination among researchers, planners, and public officials. It has been as much a political exercise as it has been an information system study. Active interdisciplinary cooperation in data collection and utilization within the University community has taken place. MLMIS staff has worked closely with government agencies to standardize data collection and storage techniques. There has been a sharing of personnel between the University and governmental agencies. Many times this resulted in permanent state employment. In addition, there are regular meetings between University faculty and users (or potential users) from state agencies. Working with these officials has proved to be a fruitful experience. Now the MLMIS personnel is more knowledgeable in the types of information needed and the types of questions asked of the information. Therefore the contents of the MLMIS and computer access systems can be better designed to meet specific user requests. The ultimate goal of the MLMIS, of course, is not simply one of data accumulation. Rather, its goal is to provide pertinent information, and, in turn, to improve the quality of public and private decisions affecting the environment.

References Cited in Text

Borchert, J. R., G. W. Orning, J. Stinchfield, and L. Maki, 1970, Minnesota's lakeshore, resources, development, policy needs: Department of Geography and the Center for Urban and Regional Affairs, University of Minnesota.

Borchert, J. R., G. W. Orning, J. Stinchfield, D. Pederson, L. Maki, 1970, Minnesota's lakeshore, statistical summary: Department of Geography and the Center for Urban and Regional Affairs, University of Minnesota.

Brown, D., M. Meyer, J. Ulliman, R. Eller, J. Gamble, S. Prestin, and D. Trippler, 1973, ERTS-1 application to Minnesota land use mapping: Report Number Three, MLMIS.

Denenberg, S. A., C. C. Corbin, and P. A. Alsberg, NARIS, A natural resource information system: Center for Advanced Computation, University of Illinois, Urbana.

Gilbert, W., 1973, Minnesota land use laws: a classification of statutory powers: Report Number Five, MLMIS.

MLMIS, 1971, State of Minnesota land use map, 1969: Center for Urban and Regional Affairs and Minnesota State Planning Agency.

MLMIS and Center for Urban and Regional Affairs, 1972, Rapid analysis fiscal tool, Annual Report for 1971-72: University of Minnesota.

Minnesota State Planning Agency, 1968, A state land inventory: Minnesota State Planning Agency, St. Paul.

Orning, G. W., 1967, The process of lakeshore development in Crow Wing County: Unpublished M.A. thesis, Department of Geography, University of Minnesota.

Orning, G. W. and L. Maki, 1972, Land management information in northwest Minnesota--the beginning of a statewide system: Report Number One, MLMIS.

Robinette, A., 1971, Land use prototype study, Empire Township, Dakota County, Minnesota: School of Landscape Architecture, University of Minnesota.

Rusch (Streed), P. G., J. R. Borchert, and G. W. Orning, 1972, Land for development adjoining northern Minnesota's national recreational corridor: Report Number Two, MLMIS.

Swanson, R. A., 1969, The land use and natural resources inventory of New York State: Office of Planning Coordination, New York State, Albany.

Wietecki, K., and G. W. Orning, 1973, Lakes in Ramsey County, recreational resource, use, policy implications: Report Number Four, MLMIS.

XX Information requests should be directed to Mr. G. W. Orning, director of MLMIS or Dr. K. Kozar, Systems Director.

TABLE 1--CHARACTERS OF PARCEL IDENTIFICATION

Type of Location	Number of Characters	Example
County	2	31
Township	3	055
Range	3	262
Section	2	36
Quarter-Quarter Section (forty-acre parcel)	2	31
Government Lot*	2	09

* All parcels which are not exactly forty acres were originally surveyed in size by the government and called government lots. Ownership is not implied.

\+ The last character in range serial is a directional code.

TABLE 2--A LIST OF HARDWARE EMPLOYED IN THE MLMIS

CDC 6600 COMPUTER

65 k words, extended core storage, 841 disk drives, eight seven track tape drives

CDC 3200 COMPUTER

32 k words, 854 disk drives, three seven track tape drives, CRT controller

CDC 211, CRT (CATHODE RAY TUBE)

Alpha-numerical model with 50 characters horizontally and 20 vertically

CALCOMP PLOTTER 563

Drum with 11" and 32" paper

INTERPRETATION SYSTEMS VP-8 IMAGE ANALYZER

(for analyzing ERTS-1 imagery)

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

SISTEMAS AVANZADOS DE GEO-INFORMATICA EN PLANEACION

COMPUTACION GRAFICA

APENDICE BIBLIOGRAFICO

JULIO, 1978.

Palacio de Minería Calle de Tacuba 5, primer piso. México 1, D. F.

Comput. & Graphics, Vol. 2, pp. 179–193. Pergamon Press, 1977. Printed in Great Britain

RASTER SCAN APPROACHES TO COMPUTER GRAPHICS

NICHOLAS NEGROPONTE
Architecture Machine Group, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

(*Received* 22 *January* 1976)

Abstract—This text reports on conclusions derived from the building and use of the 85†, an elaborate raster scan display system for computer graphics applications. The titled plurality is justified by a generality and by a superset of degrees of freedom in the specific device, all of which no one application warrants and all of which have immense programming implications. The underlying and now generally accepted assumption is that the future of computer graphics is in raster scan technologies. This paper concludes that future implementations ought to be primarily mass memories, rarely scan conversion techniques, and never run-length lists.

INTRODUCTION

History

The origins of computer graphics are in line-drawing machines. A picture was an assemblage of contiguous lines, composed of endpoints connected in sequence by visible or invisible lines structured in a display list. The display list (see Fig. 1), with or without subroutines, served the dual purpose of being both what the computer program messaged and what the display processor followed. Thus the inception of computer graphics followed a random access paradigm.

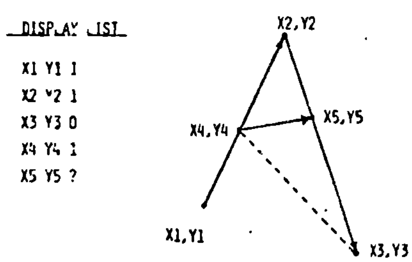


Fig. 1.

Interactive computer graphics was hatched with SKETCHPAD[1], a well-studied and less well-copied system of the earlier sixties (see Fig. 2). It launched a series of commerical efforts to embody a congenial man–machine interaction[2], particularly for applications to computer-aided design[3], to the extreme that computer graphics and computer-aided design became unintentionally synonymous. These enterprises had two bothersome characteristics in common: the hardware was expensive and the software usurped great amounts of computer time (incompatible with the then emerging notions of time-sharing).

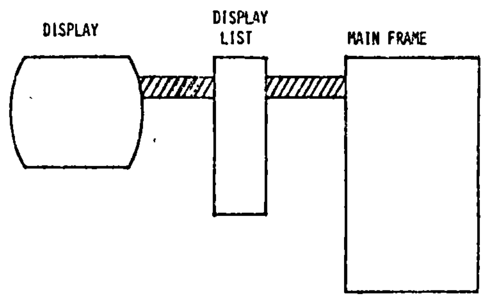


Fig. 2.

In the late sixties the storage tube emerged as a reconciliation[4], not requiring a display list, effectively a volatile plotter. The programming consequences were large, given that the picture description could be kept or forgotten in-line and the program was not at the mercy of bandwidth and buffer size (see Fig. 3). The drawbacks included slow drawing speeds and inability to erase locally, which precluded dynamic graphics (for which cost-effectiveness had yet to be proven). Shortcomings and advances aside, the storage tube still maintained a line-oriented, random-access genre of picture making. It was followed by the intelligent terminal approach of the same principle (see Fig. 4)[5, 6].

By the early seventies, there were two hardware developments, leading to a variety of actual implementations and an assortment of wild predictions: the mini-computer and mass memories. Both continue to validate the predictions of ever-increasing speed and ever-

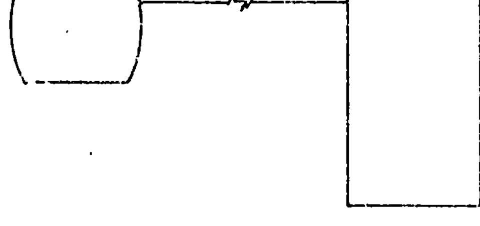


Fig. 3.

†The 85 was designed and built chiefly between March 1973 and December 1974, during which time the work was sponsored in part by the Office of Naval Research, grant No. N00014-67-A-0204-0074, and in part by the IBM Corporation. Currently, applications are being supported by the Division of Computer Research of the National Science Foundation and the Undergraduate Research Opportunities Program of the Massachusetts Institute of Technology.

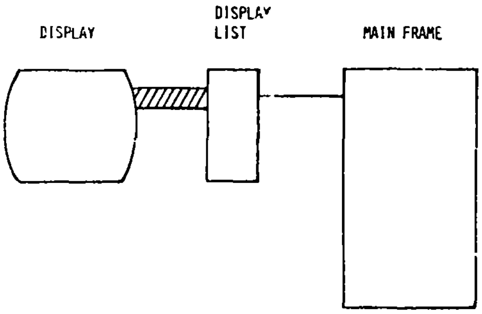


Fig. 4.

decreasing costs. One thing that did not share these orders-of-magnitude cost drops is the underpopulated, high-voltage, random-access display device. Consequently, the computer graphics community is turning cautiously toward the ubiquitous television set.

Alphanumeric terminal builders were the first to appreciate the advantages of raster scan technologies. However, their application did not bear the anomalies of full graphics. Instead, developments at Adage[7], Xerox Corporation[8], and MIT[9] launched some of the first efforts to understand and overcome the peculiarities of picture making and picture description in a raster scan device.

Three approaches

Three approaches to raster scan have emerged: on-the-fly scan conversion, run-length lists, and n-bit-per-point veridical memories. Each has programming implications and cost ramifications; this paper will dwell primarily on veridical memories.

On-the-fly scan conversion is in the spirit of a display list and accordingly affords dynamic displays. Picture elements are ordered, depending on the algorithm, from top to bottom, in such a way that hardware or micro-coded routines can determine, on a per-scan-line basis, where and when elements intersect a particular horizontal line through the picture. Early examples of this technique can be found in Xerox equipment and most dramatically in the Navy's specification and final contract for PROTEUS, which scan-converts and merges on-the-fly: straight lines, conics, and alphanumerics

Proponents of this approach argue that the logic is cheap enough, that mass memory is unnecessarily expensive (and will not drop in cost very fast), and that the dynamics are suited to their application. Without contesting these positions, one might point out that the scan conversion approach does suffer from a limitation in picture complexity, frequently determined by the maximum number of intersections with any given scan line. Also, proponents of this method disqualify the amenities of pictoral complexity in gray tone and color.

Run-length lists (Fig. 5) are an encoding technique for describing an area without having to store n-bits-per-point[10]. The assumption is that the picture will be composed of enough areas of constant tonality so that the encoding overhead and per-line description require far less computer memory than would be necessary to store every point. While this approach is far more sympathetic

1 WHITE 32 BLACK 67 WHITE
60 BLACK 40 WHITE
4 WHITE 64 BLACK 32 WHITE
6 WHITE 66 BLACK 28 WHITE
8 WHITE 66 BLACK 26 WHITE
10 WHITE 66 BLACK 24 WHITE
11 WHITE 67 BLACK 22 WHITE
11 WHITE 67 BLACK 22 WHITE
12 WHITE 68 BLACK 20 WHITE
12 WHITE 32 BLACK 10 WHITE 28 BLACK 18 WHITE
13 WHITE 27 BLACK 18 WHITE 26 BLACK 16 WHITE

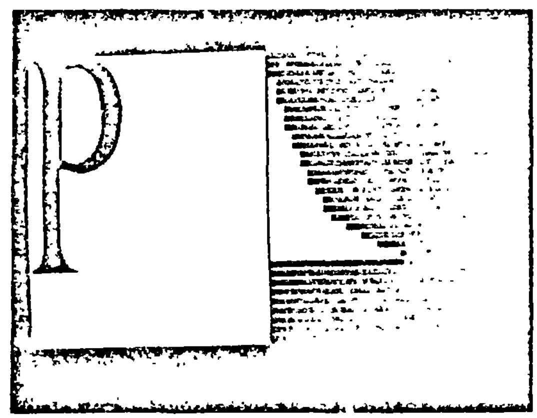


Fig. 5.

to areas and tone, it appears to be (to this author) the unhappy medium between the first and third approach and, of course, unamenable to picture processing. It affords neither the dynamics of scan conversion nor the complexities of mass-memory techniques.

The third approach is that of a mass memory, where, for every dot on the screen, there are some number of bits of memory (including zero). In the simplest sense, one can consider a 525-line television screen as a two-dimensional array of single bits 700 wide, because of the aspect ratio, and 525 high, so that, say, 1 is on/white and 0 is off/black [see Fig. 6].

Given that both the computer and the video display processor can provide access to this array, we effectively have a storage tube with local erase. Elaborating slightly, we can provide more than one bit per point to achieve gray tone or color. Ultimately, we can make the mappings of areas and tones programmable and incredibly variable.

This method was first implemented cost-effectively with disks and shift registers. However, this paper concentrates on its embodiment in random access memory, in fact shared with the computer. Critics of this method argue against it on the basis of cost and of the difficulties of handling dynamic images. The hardship of dynamics is treated later in this paper; costwise, consider that today, at $.001 per bit, 16 bits-per-point (64 k tones) on standard television represent approx. $5,000. Can't we assume, however, that this will drop at least an order of magnitude?

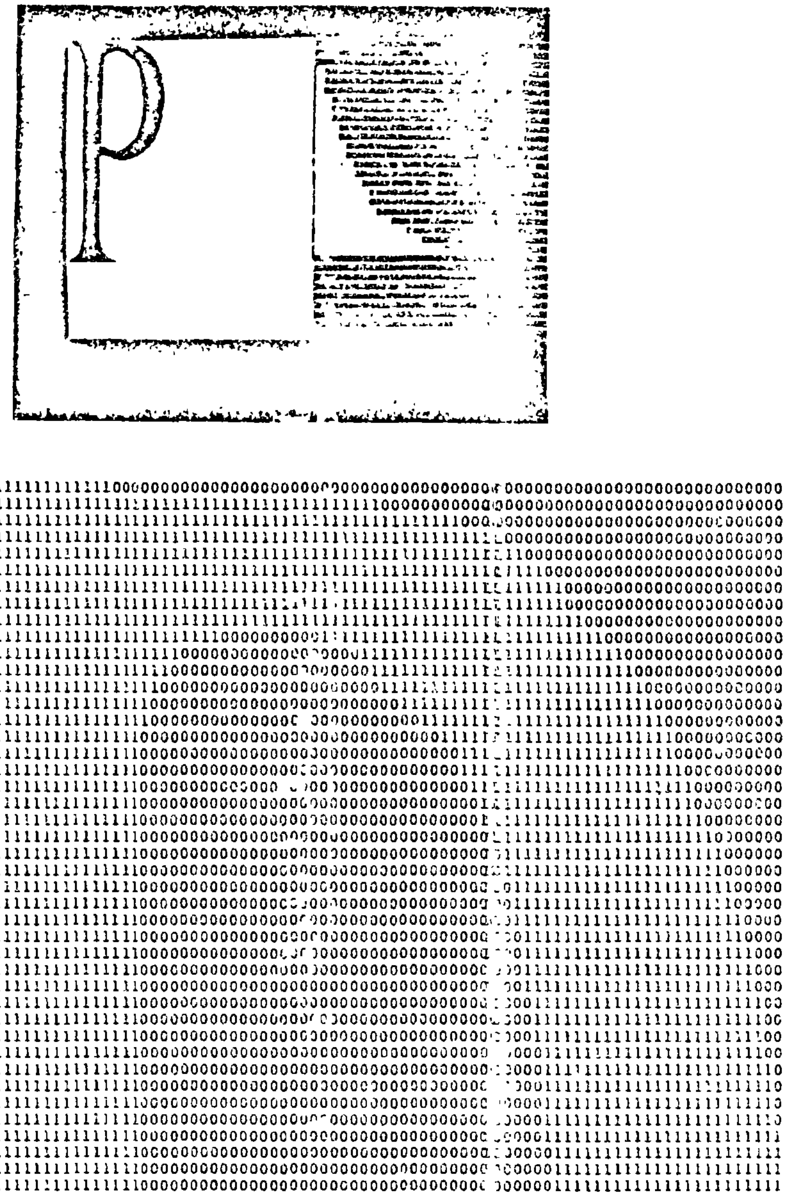


Fig. 6.

Specific origins of the 85

Given that the following sections generalize from experiments with a specific piece of hardware, the 85 (whose name is culled from its Interdata processor), this section shares the evolution of this project. While apparently circumstantial, the original needs for, and side-stepping into, raster scan techniques are indicative of future programming implications. The current machine epitomizes a transition from sequential to positional representations, in this case, in recognition and manipulation of hand-drawn sketches.

Since 1970 the Architecture Machine Group at MIT has been concerned with on-line recognition of hand drawings[11, 12], which would enable a designer to interact with computer aids at the early stages of design, when his equivocations have the graphic manifestations to which we are accustomed in sketching. In contrast to the rubber-band line vernacular of SKETCHPAD, our interest was in exploiting the freewheeling language of drawing, forcing the machine to make inferences about what the designer meant as opposed to requiring the designer to stylize his representation. In the crudest sense, the problem is one of data compression: given 400 x–y–z coordinates/s, map them into a simple representation.

Observe that both before and after the compression, the data is purely sequential (see Fig. 7). Only once intersections and latching are resolved, do the data assume the two dimensions of a planar graph. Let's consider only intersections (as latching is a much harder problem) where we would like to resolve hundreds of lines that intersect, with no reason to believe that the actual points of intersection are sampled data. The combinatorial problem is immense (greater than exponential and less than factorial). However, if we simultaneously stored a fine bit map on a fixed head disk, for

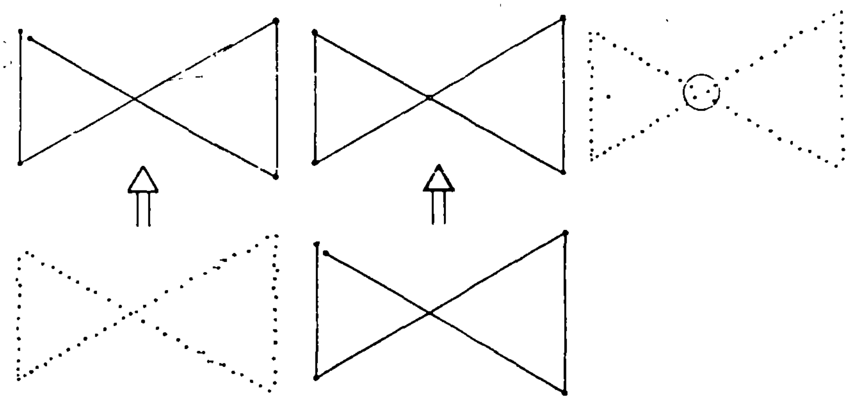


Fig. 7.

example, we could interpolate and deposit contiguous bits as we draw our lines or as our algorithms find them. In this fashion, if we bump upon a bit already on, we know we are intersecting (see Fig. 8), and the coordinates (at least) are obtained, in some sense, for nothing.

Such a mechanism was added to the mini-computer used for our experiments in sketch recognition, and all sorts of windowing and zooming facilities were developed for this bit map representation (Fig. 9). Only after completing this disk system (in fall 1972), was it conceived that the positional representation could reside in the computer's memory and, at the same time, drive a bit-per-point display. Hence our beginnings with raster scan were in a position-oriented representation of a sketch and its mappings. The display and data structure were to be synonymous.

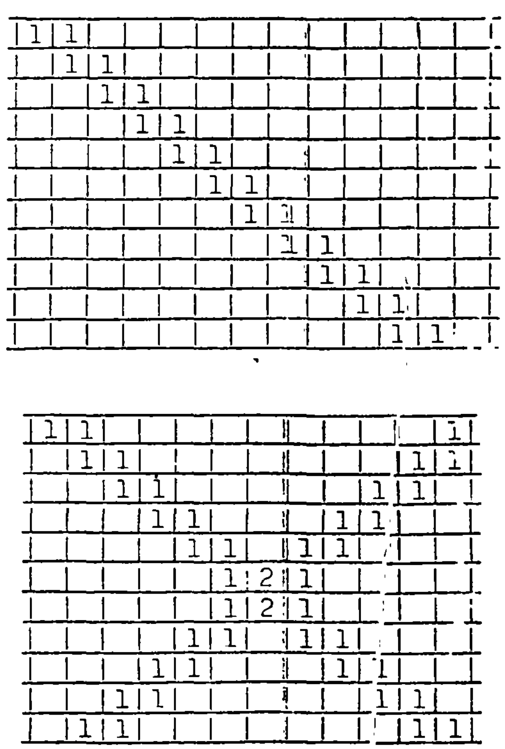


Fig. 8.

HARDWARE

Sharing mass memory

While we can argue that mass memories will become cheaper and cheaper, it is hard to pursue the line that they should be unequivocally devoted to the display and treated simply as a buffer into which one writes, or out of which one reads (a position one is locked into with shift registers). A natural tendency is to want to share the memory and, for example, in the extreme case when the display is off, to have the memory available for program execution (Fig. 13). In our previous example of 16 bits-per-point on standard American television, such a mass memory is a formidable 700 k bytes.

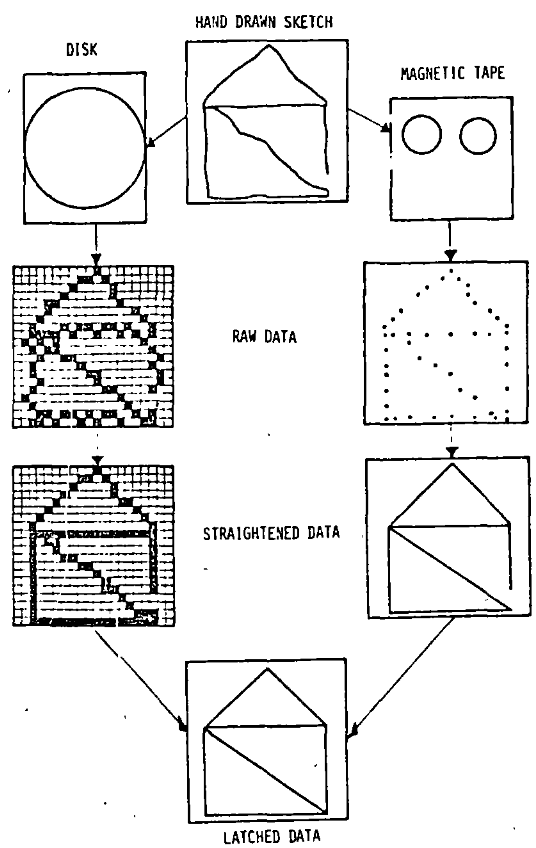


Fig. 9.

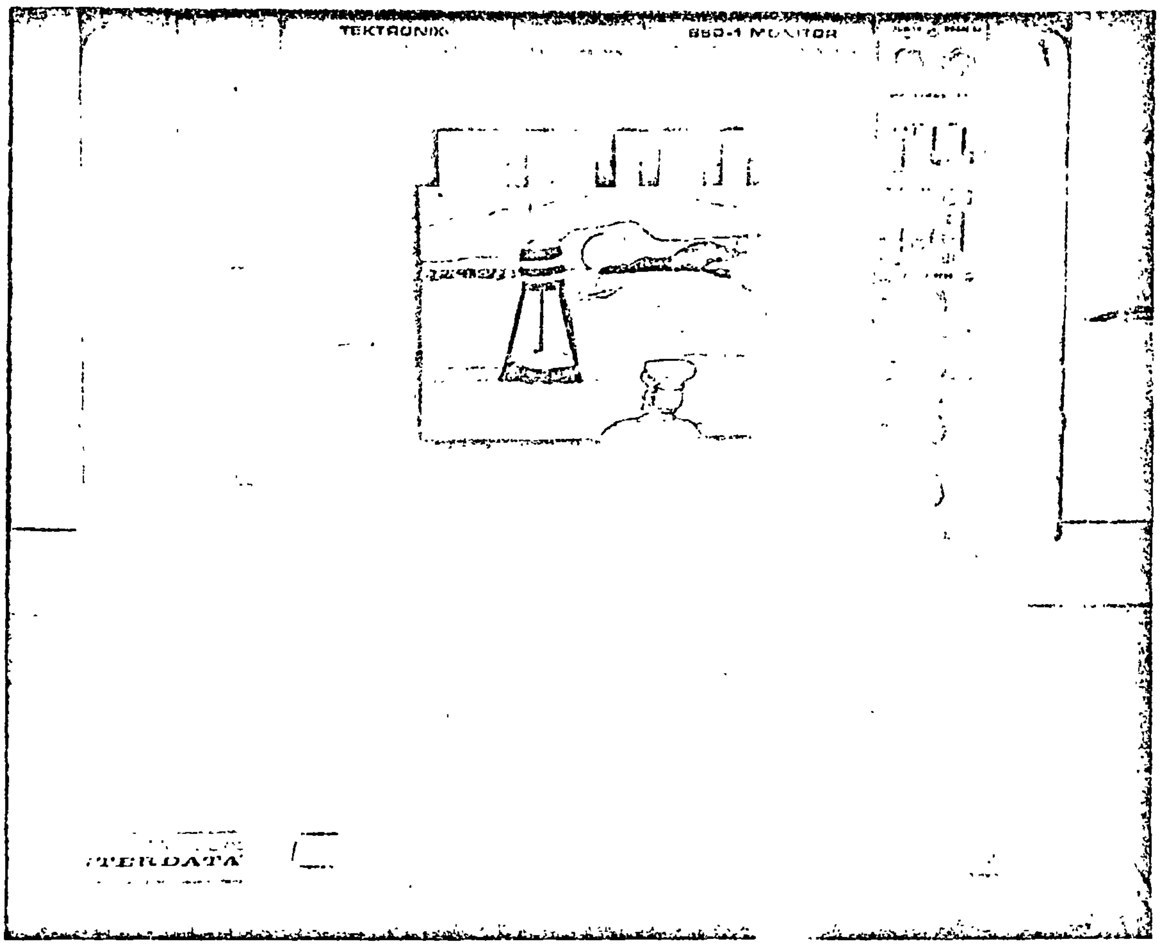


Fig. 12. The 85 used for painting.

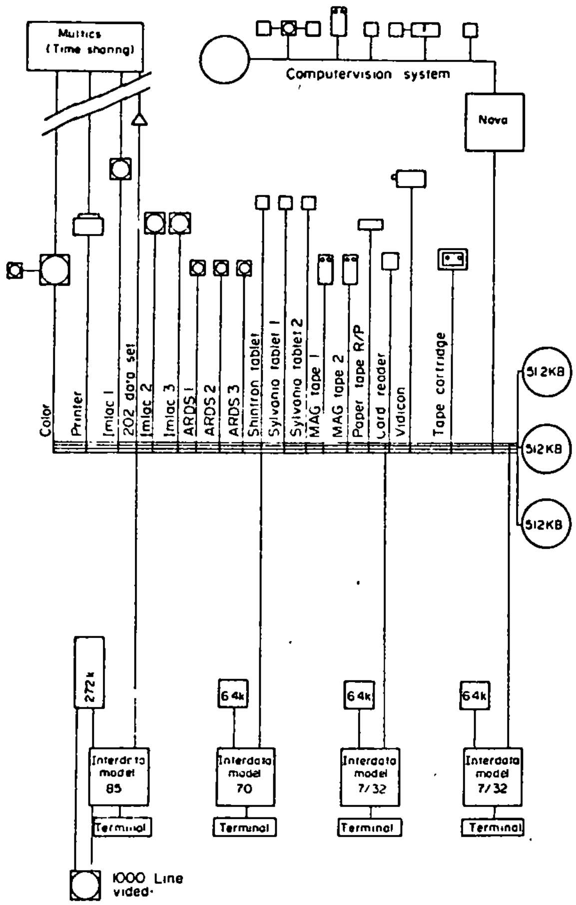


Fig. 10. The Architecture Machine, 1976.

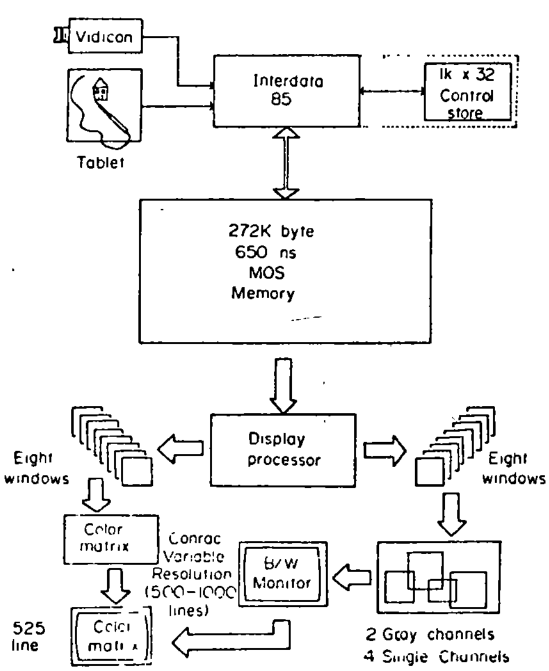


Fig. 11. The Model 85.

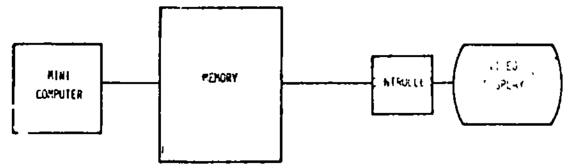


Fig. 13.

But 525 lines is not enough. One immediately finds reason to consider 1000 lines (and hope for more), which multiplies all our previous figures by 4. Conrac fortunately makes a variable-line television (the RQF4), which eliminates the need to commit oneself to a specific resolution and memory consumption (Fig. 14). By synching on the programmable signal, the display can vary its resolution. In the simplest case of single bit-per-point applications, this allows for a free-moving boundary between display space and program execution, by virtue of a variable (programmable) resolution.

For example, a FORTRAN programmer may choose to use a two-dimensional array for his picture data in a manner that allows him to make that chuck of core his display space at whatever resolution that fills the screen. But, during editing, compiling, and loading, he may choose to trade off resolution for speed (through core-residency of one thing or another) and then seek high-resolution pictures only during the execution of his program.

The same kind of sharing can be achieved by trading off the number of bits-per-point. However, unlike variable-line television, this has a crucial and immutable impact on memory organization. This is because adding or subtracting bits of gray level pre-supposes that they are not contiguous in core, except in the crude implementation of committing to 4-, 8-, and 16-bit modes. If the user wants 5 bits of gray for a photograph in one application and 6 in another, we presume that the system should, once again, have the opportunity to move boundaries between data and program execution. But this desire commits the hardware to a scheme of planning.

Planes and priorities

The memory is organized in planes, in the literal sense of animation cells. Planes can be assembled into multiple bits of gray or color (Fig. 15). Although a single point on the screen might have three bits of gray, those three bits are *not* contiguous bits of memory.

At first this may seem unduly confusing for the systems programmer. At second glance, its advantages outweigh the overhead. The last section of this paper covers applications where color separations, for example, rely upon and justify this feature. Here we will consider only the picture-making properties which come from the additional elaborations afforded by not requiring that planes be of the same size, proportion, or position. The current machine allows for twelve planes, eight of which have been implemented.

The planes are used as single-bit windows or grouped to

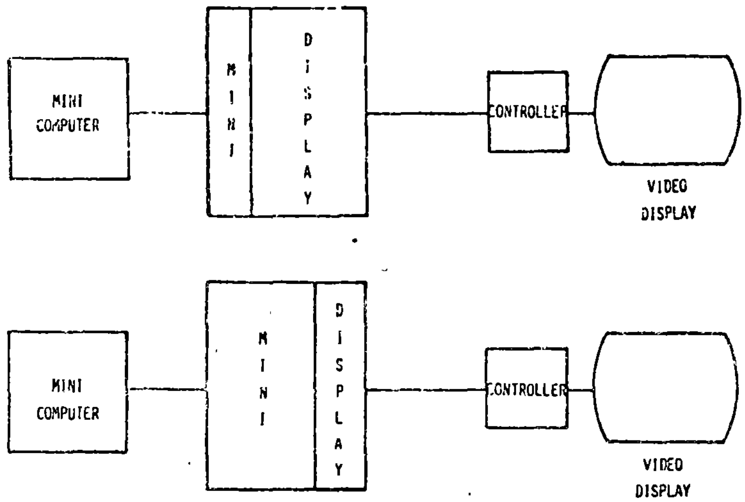


Fig. 14.

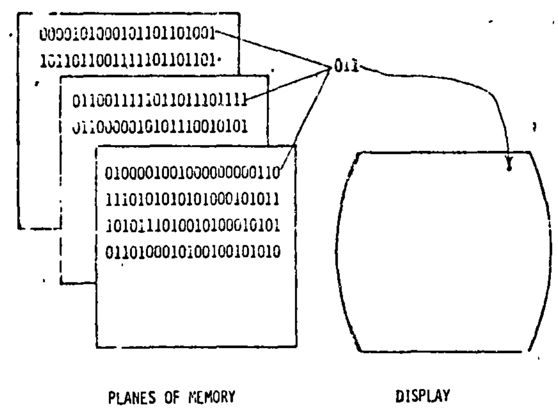


Fig. 15.

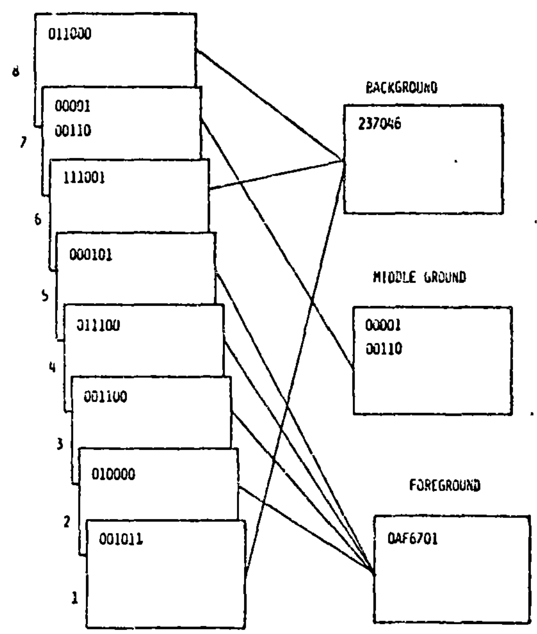


Fig. 16.

form gray tones or to matrix color. Consequently, eight planes might be arranged so that planes one, six, and eight are three bits of gray for the background; two, three, four, and five are four bits of gray for the foreground; and seven is a single bit of backline drawing in the middle ground (Fig. 16). Obviously, these need to be arranged according to programmable priorities (Fig. 17). In the event that the foreground is a photograph of an oak tree, the back and middle grounds appear only between the leaves, twigs and branches. If the oak tree for some reason is translated across the screen, the background scenes automatically appear and disappear through the foliage. In another example, the foreground might be a line drawing, so that sketches or computed vectors override and overlay a photographic background.

To add confusion to complexity, planes need not be of the same size, proportion, (Fig. 18), or overlap, and the priorities may be set as a function of a gray value, not just fully on or fully off. The latter feature has no video, photographic, or visual counterpart; the effects are unique to the notion of gray-tone priorities.

An almost degenerate application of planes would be to consider each of the eight planes as single-bit line drawings. I mention this because early attempts at using raster scan techniques for computer graphics immediately stumbled on the problem of erasure in dynamics and in line subtraction (Fig. 19). With only a single plane, a moving spot would entail a process of bit writing, substracting, offsetting, rewriting, etc., leaving a wake of whiteness. Cautious but tedious inspection of the neighborhood could allow for restoring the picture in some applications. A variable-size plane achieves the desired result most easily. That which is to be translated is moved into an auxiliary plane and

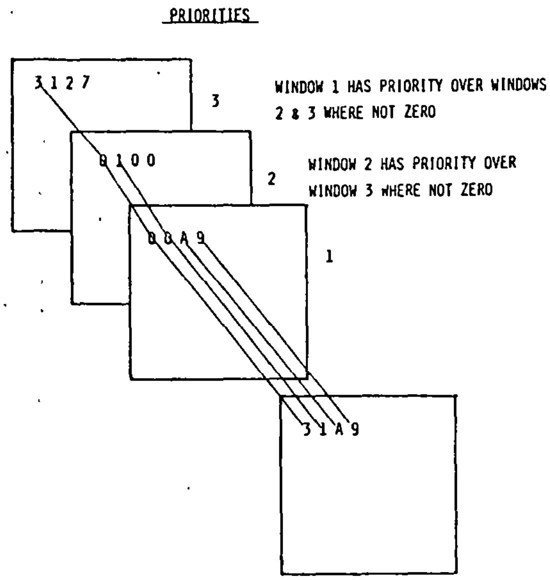


Fig. 17.

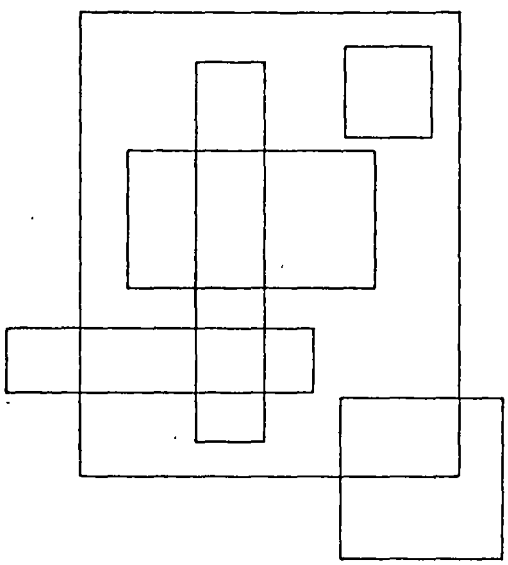


Fig. 18.

translated either by moving the plane as a whole or the object within it, while letting the priority scheme manage the background (now left intact).

Processor characteristics

The most obvious processor characteristic is addressability. Mini-computers have been inherently 16-bit mechanisms limited to 64 k direct address spaces, unable to capitalize upon mass memory for program execution. While indirect addressing can achieve data manipulation in mass memory (as the 85 has done), it makes program execution prohibitively cumbersome. Currently, Interdata

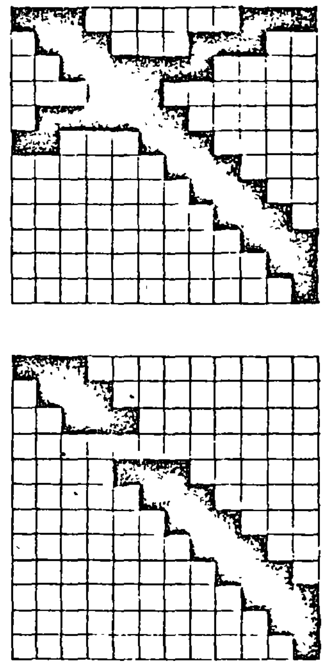


Fig. 19.

has remedied this with their 7/32 and 8/32 mini-computers. Regrettably, it was developed too late for this project.

A second characteristic is control store. The mini-computer we actually used, the Interdata 85, has 4 k bytes of writable control store (90-ns bi-polar memory). While experimentation has not explored this problem thoroughly, it is obvious that certain functions require extreme speeds, in order to be executed like other instructions.

Consider, for example, the drawing of a horizontal line. On a 1024-by-1024-bit plane, this means writing 64 contiguous words with hexadecimal FFFF. In contrast, to draw a vertical line, one must write a single bit into 1024 non-contiguous words (see Fig. 20). A sloping line makes the contiguity problem worse. While it is not difficult to write a program to compute the locations, it is slow in execution and hence more appropriately micro-coded.

Additionally, micro-code has proven fast enough to do limited scan conversion in any one of the planes, making the advantages of that approach compatible with this mass memory method.

SOFTWARE

Coordinate systems

Coordinate systems in computer graphics are caricatures of variety. Each device has its own protocol, having,

```
00000000 00000000 00000000 00000000 00000000 00000000
00000000 00000000 00000000 00000000 00000000 00000000
11111111 11111111 11111111 11111111 11111111 11111111
00010000 00010000 00000000 00000000 00010000 00000000
00010000 00010000 00000000 00000000 00010000 00000000
00010000 00010000 00000000 00000000 00010000 00000000
00010000 00010000 00000000 00000000 00010000 00000000
```

Fig. 20.

for example, its origin at the center or low left. The spatial increments vary and are by definition a function of the resolution of the device. The systems programmer worries about the transformations necessary to make each mechanism appear the same, for the sake of device independence at the applications programming level (a perpetual but vacuous aspiration of the computer graphics community). Most frequently the similarity criterion is to make a full screen image on one device a full screen on another, within the tolerance of different aspect ratios.

Alas, raster scan technologies have a 'natural' origin in the upper left-hand corner. They add further confusion and, in fact, question the similarity criteria of device independence. This is because of variable resolution, whether contiguous as previously described, or in programmable increments of 525, 625, 875, 1000, etc. Presumably, in the event of changing resolution, the user does not expect to see the size of his picture change (Fig. 21). In terms of bit maps, this requires a careful shuffling of bits, in the case of going to lower resolutions, and entire picture re-creation when going higher.

What enters into the question is the base coordinate system into which the data are stored and from which they are taken. At first glance, this is simply a matter of using large enough numbers, like + and −16 million, to be guaranteed that, first, you will map into even the highest resolution device, and, second, you can encompass a problem of just about any physical scale. This approach has been significant in the use of CRTs because one is not usually concerned with the size of the picture, in the sense of having to measure it. This is not true with most plotter applications, where the scale is often critical.

With the (elusive) advent of large, flat-screen displays, the distinction between plotter and CRT will fade. Projection television begins this convergence. The consequence in terms of coordinate systems will be the additional problem of display *scale*, previously a result of circumstance. Data will be most wisely stored in units relating to physical size, mapped onto the screen at a specific scale.

Shape orientation

Before raster scan displays, lines had a length and, on some displays, bore an intensity or color. Now, they additionally have an inherent width (Fig. 22) (without the disadvantage of juxtaposing parallel lines). In the most general sense, a line is now a special instance of an arbitrary blob, i.e. a rectangle (Figs. 23 and 24). The result is that traditional programming strategies and higher-level graphics languages lose much of their usefulness. It is more a problem of cellular automata than a problem of vectors.

When we limit ourselves to two dimensions, the descriptive problem is two-fold: that of boundary description and that of containment. Picture processing techniques offer some clue for the former, and scene analysis has worried with the latter. In computer graphics it remains a virgin problem, with the exception of a few animation exercises where image clipping is applied to planes.

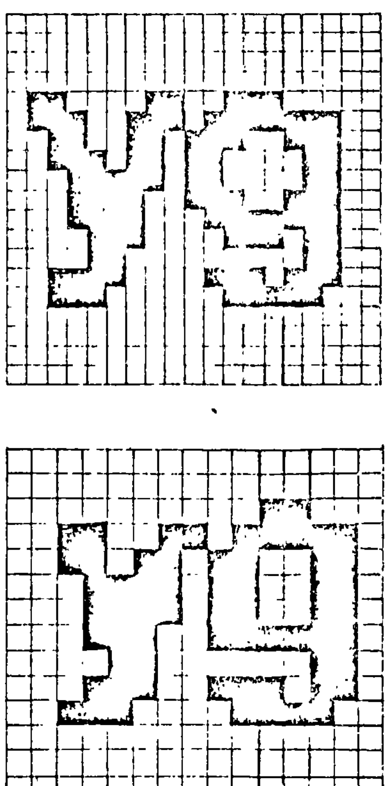


Fig. 21.

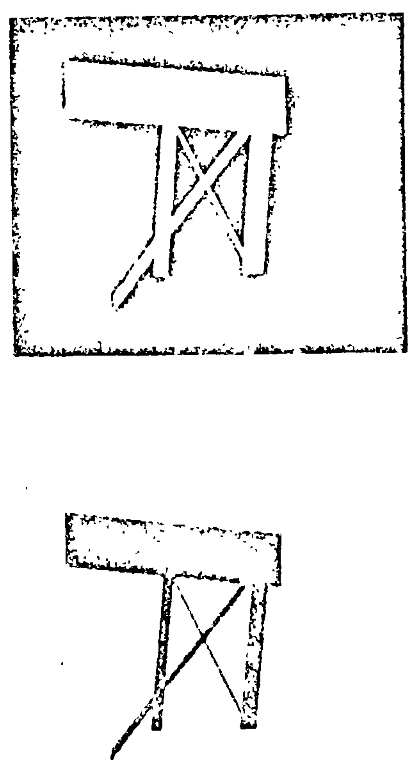
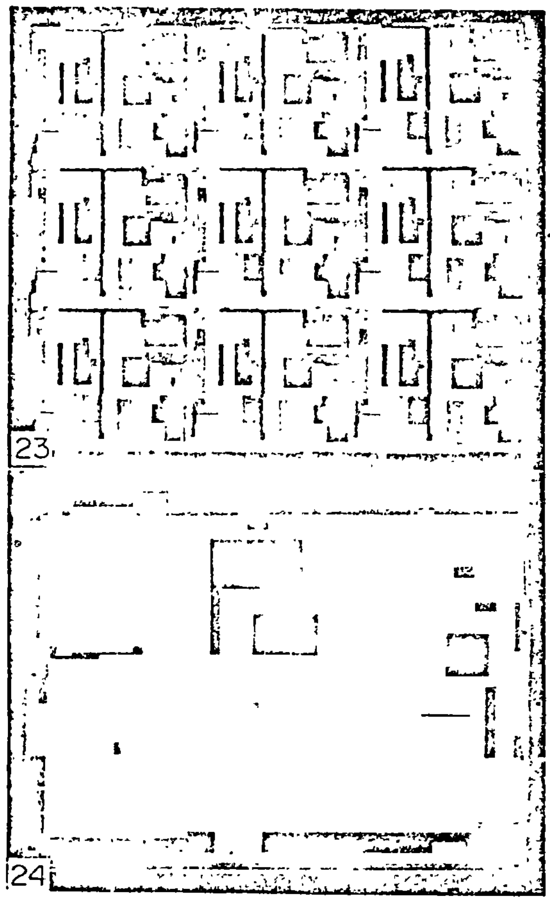


Fig. 22



Figs. 23 and 24.

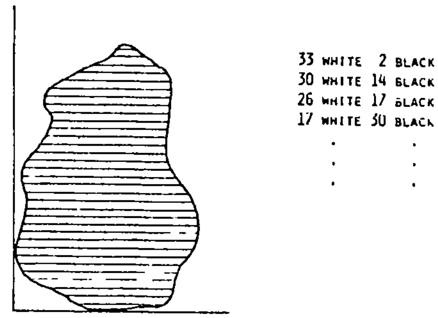


Fig. 26.

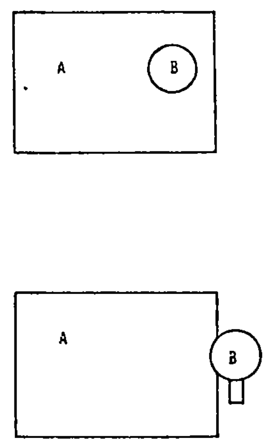


Fig. 27.

In describing the boundary of a figure one has the choice of defining its contour as a path or its shape as a filled-in area. The contour approach (Fig. 25) appears more appropriate for extremely complex figures, typical, for example, of medical and biological applications. Use of filled-in areas (Fig. 26) tends to lend itself to regular geometries with few concavities, more typical of engineering applications. In the one case the boundary is described and an algorithm fills it with a tone or color. In the other case that algorithm is the description.

Containment is the shape counterpart of linkage structures in line drawing. Just as line A and line B can share an endpoint a or be edges of a rectangle R, or part of a camshaft c, blob A may be within B, part of B, a hole for a rod, (Fig. 27) etc. Note that this association is in the problem description and not the picture description. Areas A and B may simply overlap and not have structural association.

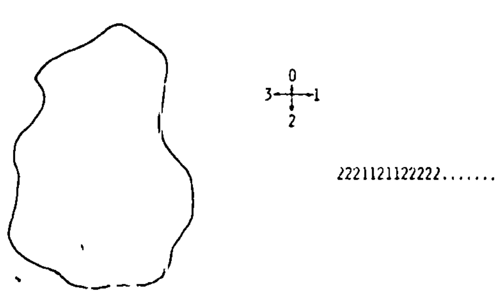


Fig. 25.

These problems are new and unresolved. This report is about their discovery, not about their solution (yet). They are further aggravated with the addition of picture processing.

Picture processing

Picture enhancement and analysis have developed into a discipline quite separate from computer graphics. Graphics enthusiasts have accepted and taken for granted the barren background of a (usually) dark green phosphor or the whiteness of a sheet of paper or television. Few applications have sought to mix drawings with photographs, whether to create richer backgrounds, or to make them manipulable by and meaningful to the machine. This is in contrast to rear projection techniques or so-called peek-a-boo tubes that can superimpose a slide, film, or video signal, picturesque but unattainable for the computer.

A raster scan approach to computer graphics provides the new opportunity to mix, manipulate, and superimpose photographs and drawings (Fig. 28). A photograph is distinguished from a shape only inasmuch as the 'filling' is no longer a single tone, color, or constant gradation, but an apparently random assortment of gray tones and possible hues. The use of photographs eliminates the previously acceptable descriptive procedure of describing

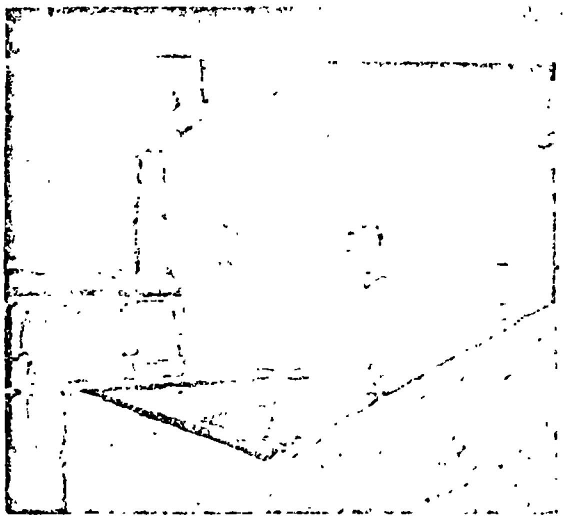


Fig. 28.

a boundary and filling it. The figure must be described in a bit-by-bit fashion, faithful to the original input. We assume that the complexity of the photograph—a tree, a face, or a country setting— is so varied that there is no merit in encoding it line by line.

The programming consequence is that we now have entities which are difficult to scale and rotate. Notwithstanding this difficulty, the potential of picture making and editing is enormous.

Color

Computer graphics has worried about the effectiveness and cost benefits of color to an extent proportional to its previous cost, but out of proportion with current techniques. In raster scan, we can presume color to be almost free.

The programming implications are not as simple. A serious gap exists between the color photograph applications and the representation of the many but limited colors of a sketch or schematic. A popular implementation for the latter is to matrix the color so that n bits per point are not committed to a specific combination of red, green, and blue. With eight planes, one has the opportunity of attaining 256 colors, which, with a matrix scheme, may be 256 shades of pink or a course section through the entire spectrum (see Fig. 29).

Picture processing applications (for example, LANDSAT data) tend to require more colors, which, in the implementation described, mean more planes and a larger matrix box. Ultimately, the notion of planes makes less sense, and the programmer would be much happier to see, say, his five bits each of red, green, and blue as contiguous in a half word (see Fig. 30). This is a fundamental incompatibility with the planing approach for which we have lobbied and which can be resolved only with the specifics of an application. For example, problems of color separation would profit from the planing feature.

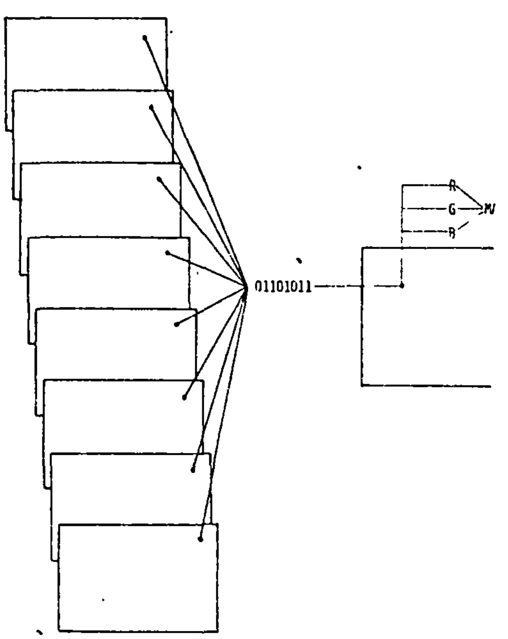


Fig. 29.

SOME APPLICATIONS

Graphic design

The concerns of a graphics designer include the typography of books, the layout of journals, directional signs for cities, and the fabrication of advertisement and symbolism. This is what we will mean by graphics in this section. It has been chosen as a sample application because of the inherent need to mix high-quality

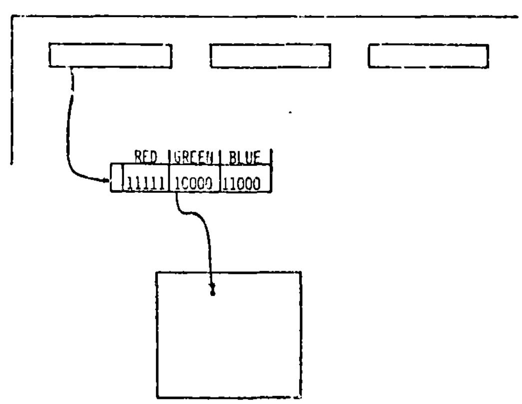


Fig. 30.

Fig. 31.

characters, drawings, and photographs (Fig. 31). Currently, in page layout, gallerys, artwork, and photographs are pasted up in a most laborious, time-consuming manner. Efforts at automation[13–15] are still in their infancy, predominantly off-line procedures. Raster scan techniques offer the opportunity to assemble camera copy on-line.

The use of raster scan displays for high-quality text was first illustrated at Xerox Corp. (Figs. 32–34, by courtesy of Xerox Corp.), using on-the-fly scan conversion techniques on an 875-line television display. Since, others have improved and implemented these techniques so that the typeface on a CRT can approximate and even emulate those to which we are accustomed in typeset print. The

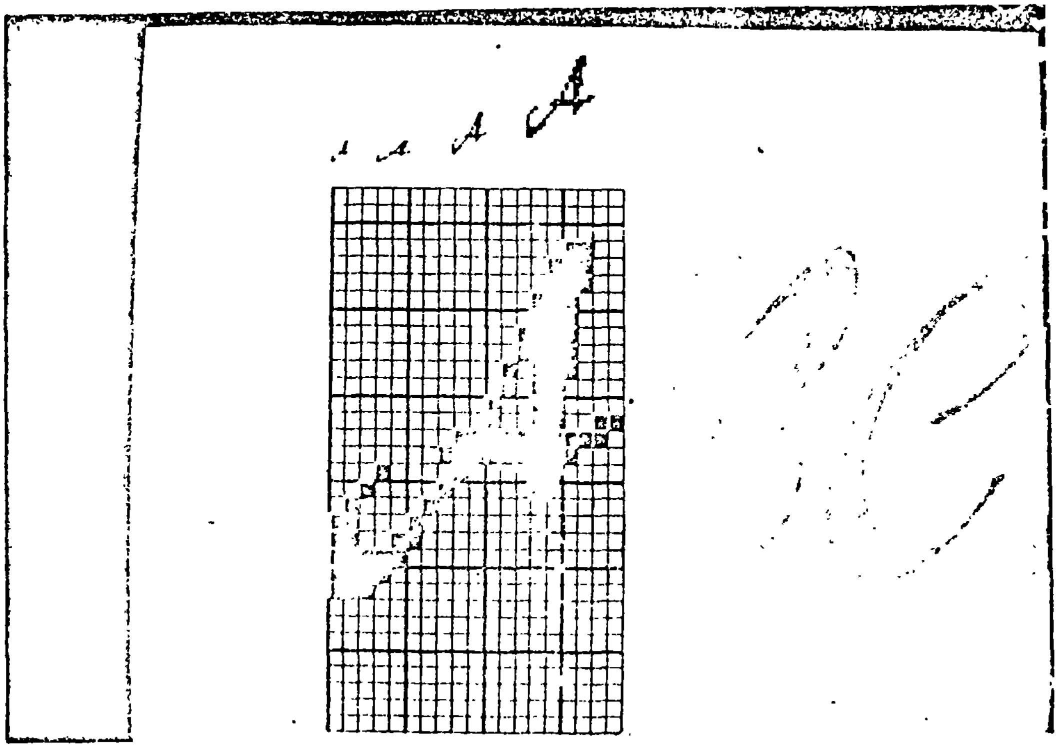
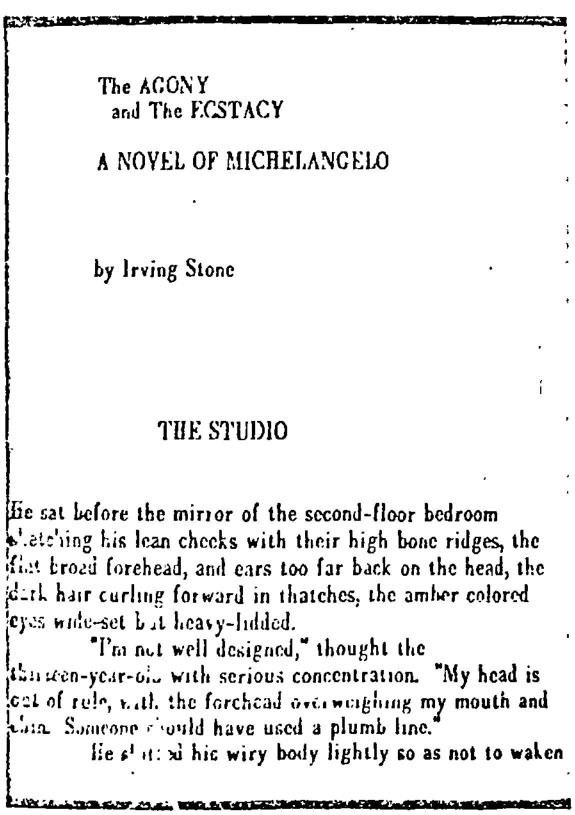


Fig. 32.



The AGONY
and The ECSTACY

A NOVEL OF MICHELANGELO

by Irving Stone

THE STUDIO

He sat before the mirror of the second-floor bedroom sketching his lean cheeks with their high bone ridges, the flat broad forehead, and ears too far back on the head, the dark hair curling forward in thatches, the amber colored eyes wide-set but heavy-lidded.

"I'm not well designed," thought the thirteen-year-old with serious concentration. "My head is cut of rule, with the forehead overweighing my mouth and chin. Someone should have used a plumb line."

He shifted his wiry body lightly so as not to waken

Fig. 33.

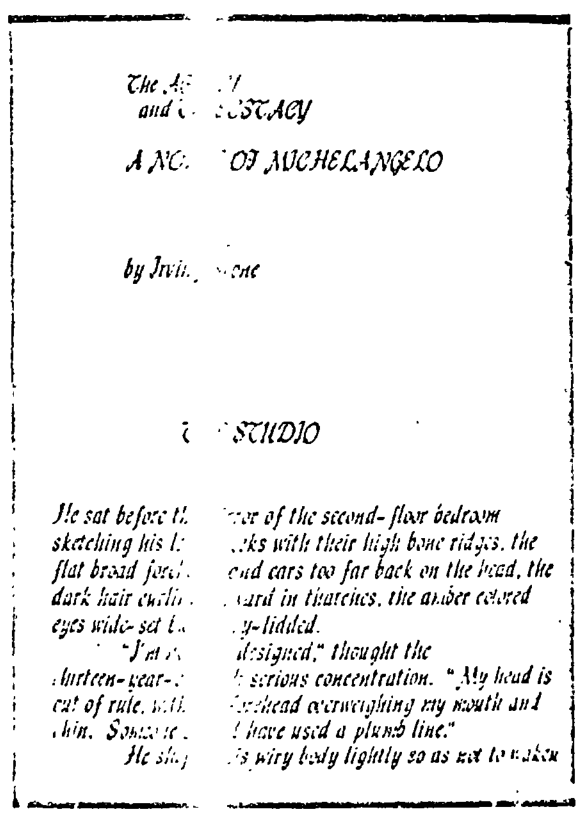


Fig 34.

original Xerox work presumed that the resolution of the display was high enough (and the screen small enough) not to warrant removing the 'jaggies' in the character description. It is true that above a certain number of lines-per-inch of video (more than 100) the sense of horizontal line indigenous to television goes away (leaving a 'smooth' background) and that with it goes much of the annoyance of jaggies in characters. Our work at MIT presumed, however, that much of the satisfaction resulted from the 'squiggly' nature of characters, because even at extreme resolutions (300 points-per-inch) a long, straight, sloping line (for example, in a 15 point Times Roman M) exhibits annoying steps.

The solution is to digitize characters on a large grid. If our Times Roman M is read in with a vidicon on a 300-by-300 single-bit map (Fig. 35), it can be reduced to any size, generating gray-tones at the edges (the percentage of 'hits' per call in the reduction), thus removing jaggies. The optical effect is one of smooth sloping edges and curves (Fig. 36).

Line drawings and shapes are more cumbersome, particularly when overlayed on backgrounds of varying colors. The problem of jaggy removal is complex and must be done in line with the creation of the image. The problems is surely the most studied aspect of raster scan display techniques[16].

In graphic design applications, photographs assume incredible importance as soon as they can be cropped and sized on-line. Currently, we postulate a two-screen system where the photograph is read and presented on one screen for cropping and then scaled for placement in the text on the other.

In graphics application the output medium remains at roadblock. Only the most expensive devices can lay out photographs, text, and line work on the same photographic or xerographic medium. Currently, video (at 1000 lines) just misses the requisite resolutions. Two thousand lines of video over a 10-in. area would satisfy most printing needs, but requires the four-fold increase of mass memory, probably exceeding even our enthusiasm about cost drops over the next two years.

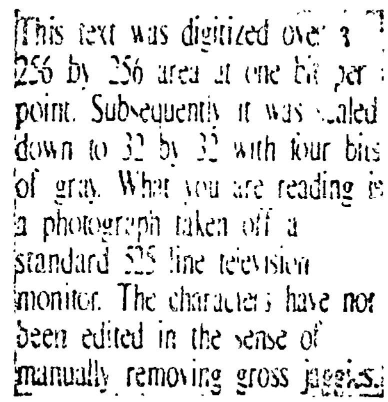


Fig. 36.

Animation

All dynamic graphics is animation in the sense of creating frames of differing content to support the illusion of movement. This section is limited to animation in the entertainment and art sense of animated films. Its distinguishing feature (which sets it apart from, for example, Computer Output Microfilm) is that it is on-line.

There are two consecutive consequences to the real-time aspect of the 85 as an animation machine. The first is the medium of input, which can be in itself graphic, in part or in whole. This permits the formation of key frames, diagrammatic description of movements, and editing of backgrounds and 'characters'. As an artistic medium, it affords the opportunity to 'paint' in time and play with effects of transparency and transfiguration.

In turn, the user can now be a different person from the one we have previously witnessed using computer animation techniques. He is much more an animator and much less a computer enthusiast. In fact, he could be a child literally fingerpainting.

Scene processing

The future television studio will be pre-dominantly, if not completely, digital. Currently, the editing of videotapes is cumbersome, to say the least. The machine

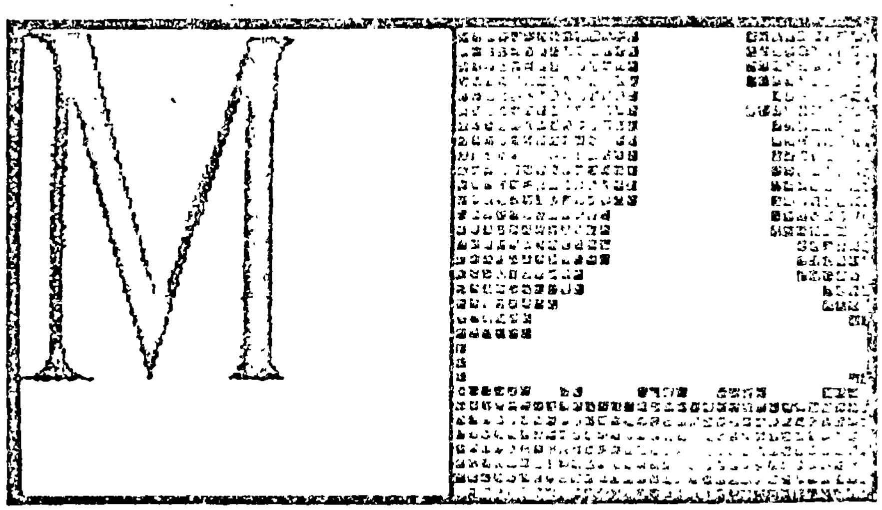


Fig. 35.

described here, and its predecessors and successors, are the beginnings of digital approaches to video.

Aside from the straightforward editing of video or photographic information, scene processing features in raster scan techniques are most valuable in their overlay upon and mixing with alphanumerics and graphics. Consider the three following sorts of examples.

An architect is interested in sketching a perspective or having the computer generate one of a proposed design. He not only wants to see the three-dimensional construct of his design, but he would like to put it in its setting, currently achieved by using abbreviated line work to depict neighboring structures and vegetation. In contrast, this brand of computer graphics would allow him to place his design in a photograph (someday a videotape) of the environment, presuming that he or the computer has worried about compatiable vantage points and focal lengths. It is not hard to imagine automated or interactive removal of the hidden scenery and replicating the foreground.

A second example might be a weather forecasting system, where satellite photographs reveal cloud formations that (let's assume) the computer can recognize, illustrate, and label, and to which it can add arrows to show their direction of movement. This kind of scenario can be postulated in any application where the photograph (or X-ray) is to be diagrammed through the results of picture-processing techniques or human intervention. This has been suggested for the on-line diagramming of football plays.

As a final example, imagine painting a photograph. At the time of writing, the notion is totally speculative and results from observing continued efforts to achieve a realism in drawing that may well be feasible with a 'knowledgeable' stylus. If you wanted to draw a tree and there was a photographic tree at hand, it could be simply placed in the scheme at the correct size and position. More exciting would be a 'likeness' of tree that could yield a variety of foliages in the same species with procedural knowledge of how to move randomly around the greens, fatten but not elongate, and so on. Or finally, imagine a brush stroke of 'brickness' that can be assembled, computing the proper perspective, perhaps with an understanding of the role and need of expansion joints!

CONCLUSIONS

It would be foolhardy to replicate the particular machine we have built. Its mainframe is outdated, its display processor needs redesigning and simplification, and the software bears the marks of circumstance and adhocism. However, as a superset of features, it continues to support a broad range of applications, exercising many approaches to raster scan. We cannot conclude with a specific recommendation for one technique vs another. Instead, we conclude with a general enthusiasm for raster scan technologies in computer graphics, based on a host of experiments.

One aspect of this enthusiasm has not been stressed enough. That is the potential of television compatibility, underntining in part the variable-line freedom of the existing paraphernalia. However, variable-line features are an excursion away from standard television which may not be worth the loss in compatibility. This concerns us because of the simple observation that there are 121,000,000 television sets in use in the U.S.A., and 57,000,000 of them are color[17].

Computers in the home, as a ubiquitous consumer product, are no longer complete fantasy. Most of us have a beginning component for the graphics appendage to this home computer. More modestly, we can view it as an 'intelligent terminal' to the (temporary) amenities of time-sharing services.

Bars, airports, student unions, and many public places already host crude television-based games that *are* (though modest) raster scan approaches to computer graphics. These will be extrapolated orders of magnitude in areas of computer-aided instruction, office automation, and, perhaps most important, the manufacturing of toys.

REFERENCES

1. I. E. Sutherland, SKETCHPAD: a man–machine graphical communiation system, MIT Lincoln Laboratory TR 296 (May 1965). Abridged version in Spring Joint Computer Conference Proceedings, 1963, Spartan, Baltimore, MD.
2. Carl Machover, Computer graphics terminals, a backward look, *Am Fedn. Information Processing Socs. Conference Proc.* 40 (1971).
3. S H. Chasen, *The Introduction of Man-Computer Graphics into the Aerospace Industry,* Fall Joint Computer Conf. Proc., Spartan, D.C. (1965).
4. R. Stotz, A new display terminal, *Computer Design Magazine,* (April 1968).
5. C. I. Johnson, Principles of interactive systems, *IBM Systems J.* 7, 147–174 (1968).
6. W. H. Ninke, A satellite display console system for a multiaccess central computer. *Proc.* 1968 *Int. Fedn. Information Processing Congress,* (Edited by A. J. H. Morrell), p. 962. North Holland, Amsterdam (1969).
7. Daniel E. Thornhill & Thomas B. Cheek, Raster-scan tube adds to flexibility and lower cost of graphic terminal, *Electronics,* 95–101 (7 February 1974).
8. A. Goldberg & A. Kay (Eds), *Smalltalk*-72 Instruction Manual. Xerox PARC (March 1976).
9. Jeffrey Entwisle, An image processing approach to computer graphics, *Proc. IEEE Conf. Computer Graphics* (July 1974)
10. Gregory M Hunter, Full-color television, from the computer, refreshed by run-length codes in main memory, Tech. Rept No. 182, Computer Science Laboratory, Princeton University (21 April 1975).
11. Nicholas Negroponte. Sketching—a computational paradigm for personalized searching, *AIA J.* (Fall 1975).
12. Nicholas Negroponte, Recent advances in sketch recognition (reprint), 1974 *Best Computer Papers,* (Edited by Isaac L. Auerbach), Petrocelli. New York (1974).
13. Peter B. Denes & Ira K. Gershfoff, An interactive system for page layout design. *Proc. ACM* 1974 Annual Conference, San Diego, CA, November 1974, p. 212–221. ACM, New York (1974).
14. Nick A. Farmer & Joseph C. Schehr, A computer-based system for input, storage, and photocomposition of graphical data, *Proc. ACM* 1974 Annual Conf. San Diego, CA, November 1974, p. 563–570. ACM, New York (1974).
15. Brian W. Kernighan & Lorinda L. Cherry, A system for typesetting mathematics, *Commun. ACM* 18, 151–157 (March 1975).
16. H. Klaasman, Some aspects of the accuracy of the approximated position of a straight line on a square grid, *Computer Graphics and Image Processing* 4, 225–235 (September 1975).
17. EIA, Electronic Industries Association, Consumer Electronics Annual Review (1975).
18. Sherman H. Boyd, Digital-to-visible character generators, *Electro-technology* 75, 77–88 (January 1965).

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

SISTEMAS AVANZADOS DE GEO-INFORMATICA EN PLANEACION

ARTICULOS TECNICOS

JULIO, 1978.

SISTEMA DE INFORMACION GEOGRAFICA

VISION GENERAL.-

El propósito de este artículo es el de señalar la finalidad de los sistemas de información geográfica, establecer un marco de referencia para clasificarlos y describir algunos sistemas típicos.

I N D I C E

DEFINICION

Un sistema de información geográfica o sistema de geoprocesamiento es todo aquel sistema de computación donde la información que se maneja tiene el atributo de posición geográfica de la entidad generadora de la información, haciendo uso de ella de forma eficiente, para relacionar la información entre sí.

Algunos ejemplos de las operaciones que se pueden realizar son:

a).- la correlación de información, donde el sistema nos puede proporcionar automáticamente los datos de las entidades vecinas, o de las que están comunicadas entre sí, - por carretera, telefono u otro medio.

b).- la generalización, mediante la cual podemos agregar la información para que esta, a un nivel menor de detalle, nos refleje las condiciones de áreas cada vez más extensas.

El objetivo de almacenar la información con su posición geográfica es el de establecerla como marco de referencia, que resulta universal. Otro marco de referencia es el del momento en el que se capta la información. Identificándole con estos dos parámetros tendremos un -- identificador que no puede ser duplicado, ya que en un momento determinado y en un lugar determinado solo puede existir una sola entidad.

La identificación precisa del lugar y la fecha permiten establecer un código que puede servir de manera unívoca para la identificación de -

recursos naturales, de infraestructura y humanos. El conocimiento de ellos es esencial en las labores administrativas y de planeación de - gran parte de las labores institucionales cuya actividad tenga una -- dispersión geográfica.

TOMA DE DECISIONES

Las causas fundamentales por las cuales se está recorriendo cada vez en mayor medida a los sistemas de información geográfica son básicamente dos:

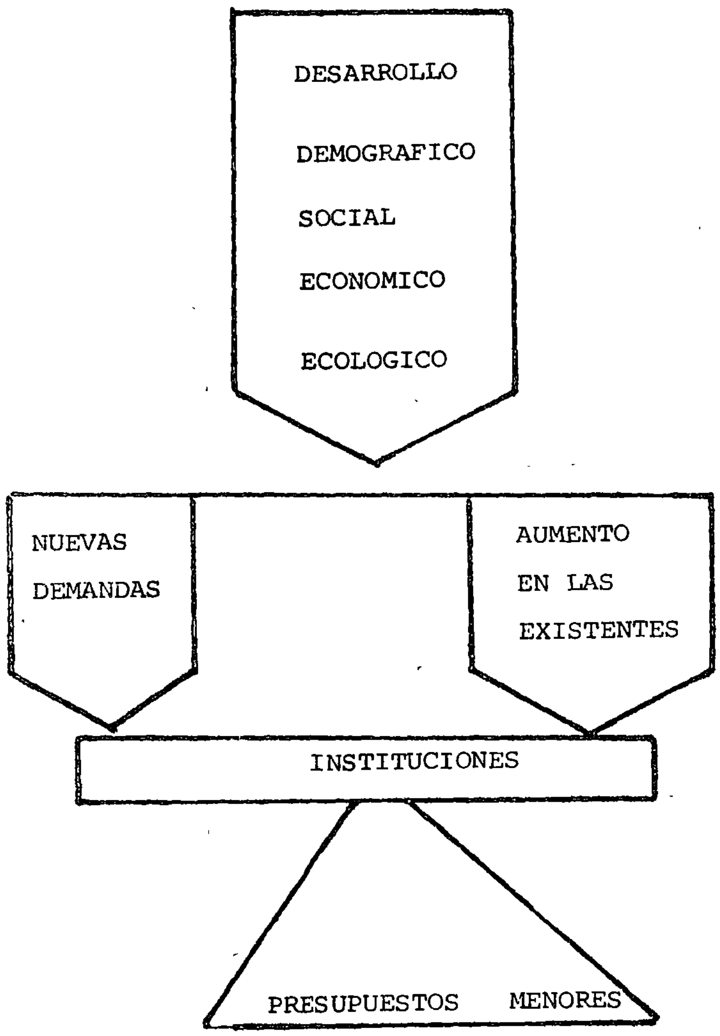
- El crecimiento en la demanda de todo tipo de servicios.
- la cada vez mayor compleja interacción entre los factores que conforman a una sociedad.

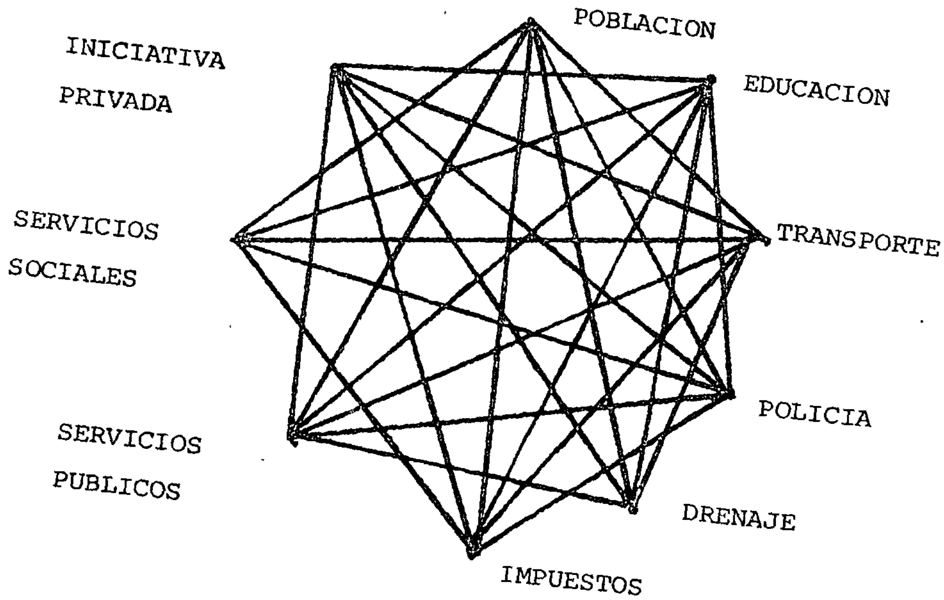
El desarrollo demográfico, económico y social de nuestros tiempos - está creando por un lado nuevas necesidades, pero también está aumentando la demanda de las ya atendidas anteriormente.

A esto se suma la reducción proporcional de los presupuestos, pues si bien éstos crecen, no lo hacen en la misma proporción de la demanda.

Es interesante señalar que para el año 2,000 se tendrá que construir un número de viviendas y servicios equivalente al construido hasta la fecha, dado que para el fin de siglo la población será el doble de la población actual.

Por otra parte, todos los factores que conforman a una sociedad interactuan entre sí y las decisiones de cualquier sector y a cualquier ni


DESARROLLO
DEMOGRAFICO
SOCIAL
ECONOMICO
ECOLOGICO
NUEVAS DEMANDAS
AUMENTO EN LAS EXISTENTES
INSTITUCIONES
PRESUPUESTOS MENORES



INTERACCION DE LAS ACTIVIDADES

vel afecten a todos los restantes.

Así por ejemplo, si se construye una industria en una población esta tendera a atraer a la población lo que hará que se requieran servicios públicos, sociales, viviendas, etc.

Un ejemplo reciente de la complejidad de esta interacción es el vivido recientemente con la construcción de la siderúgica Lázaro Cárdenas Las Truchas, donde al principio no se pudo prever con exactitud ni se pudo coordinar la interacción de las diversas dependencias que proporcionaron los servicios requeridos por la afluencia de habitantes, presentándose fenómenos de escacez, carestía y especulación.

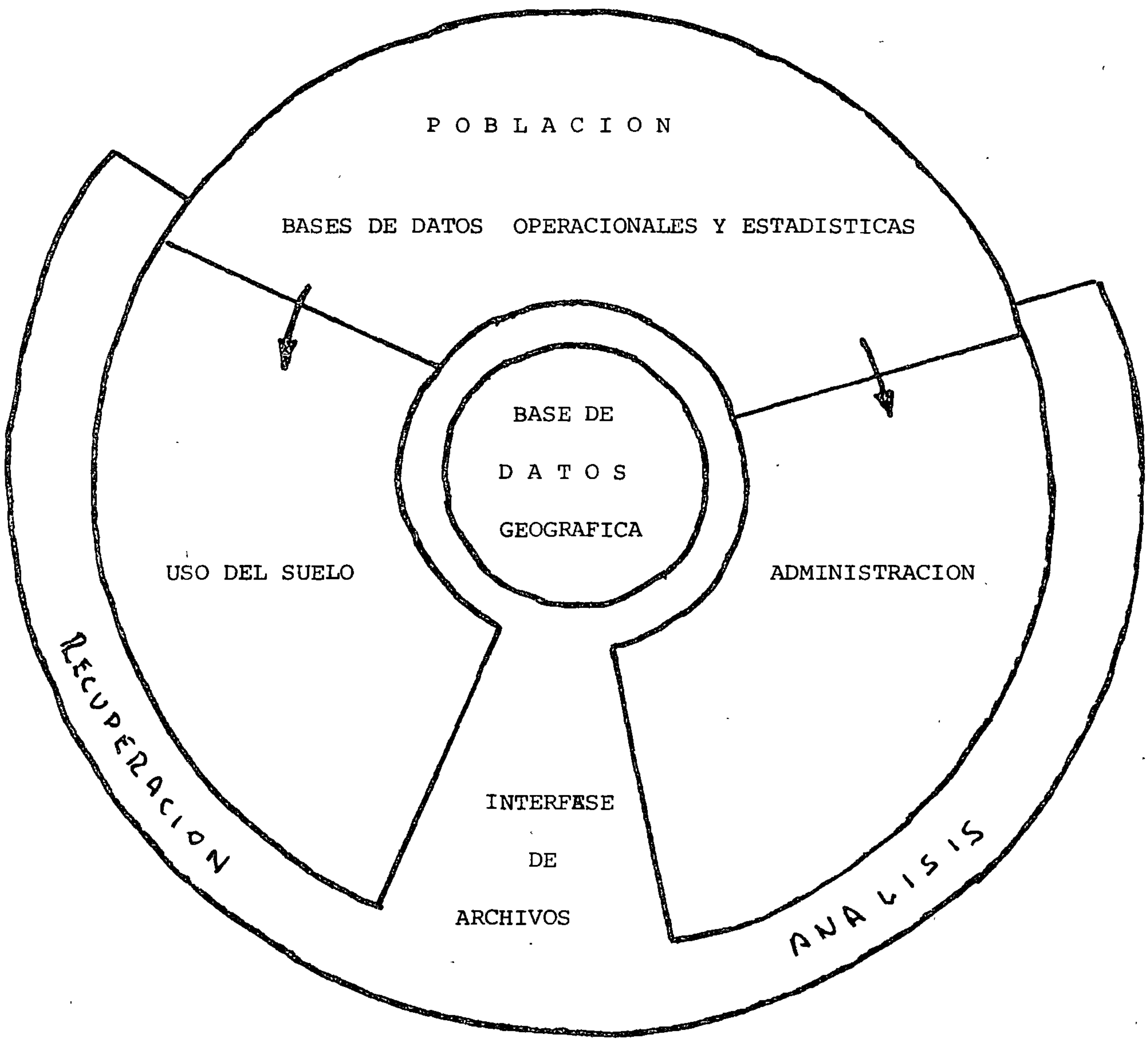
Es por estas causas que se requeria de sistemas que permitan la más efectiva y ágil planeación. Para efectuar una buena planeación se requiere de información y esta debe integrarse de manera coherente.

INTEGRACION DE INFORMACION

La mejor forma de lograr esto es a través de una base de datos geográfica que permite, a través de un lenguaje de interfase, la recuperación y el análisis de la información.

Básicamente las decisiones que respalde dicho sistema son con respecto al uso del suelo y en decisiones administrativas de las instituciones que tengan sus actividades dispersas geográficamente.

Un concepto fundamental que debe respetar esta interfase entre las -



INTEGRACION DE INFORMACION CON BASE GEOGRAFICA bases de datos es de que se deben de poder aprovechar las bases de datos operativas de las instituciones con poco o ninguna necesidad de intervención manual.

NIVELES DE INFORMACION

Por otra parte, el sistema de información geográfica debe adaptarse a los diferentes tipos de decisiones que se pueden presentar: Esto - es, decisiones operativas, administrativas, tecticas, estratégicas y políticas. Para esto debe poder agregar la inform ación desde los niveles más detallados a los niveles más generales de manera automá tica.

El objetivo de esto es el de que la información con la que se toman las decisiones políticas sea la misma que con la que se toman las de cisiones técnicas y administrativas.

En ocasiones hay una interrupción en el flujo de información y las de cisiones no son consistentes porque la información no lo es. Un ejem plo de agregación sería en relación a un marco de referencia donde - se tuviera, la división de país en estados, municipios, localidades y predios. Esta serviría para toma de decisiones políticas, estraté gicas, tacticas, administrativas y operativas para todo el país.

Para un área urbana se podrían tener los niveles de áreas metropolitana, delegación, sección, predio y facilidades.

No hay que olvidar que las divisiones a nivel nacional afectan a las regiones y que en general no existen a la fecha mecanismos adecuados

N I V E L E S

AGREGACION DE INFORMACION

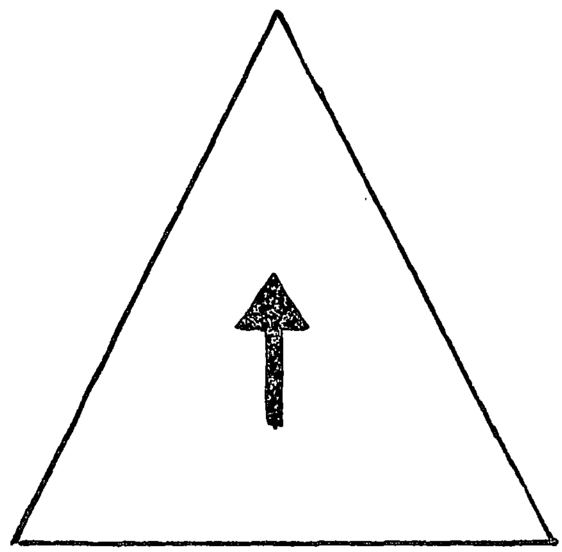
PAIS O AREA METROPOLITANA

ESTADO O DELEGACION

MUNICIPIO O SECCION

LOCALIDAD O MANZANA

PREDIO



TOMA DE DECISIONES

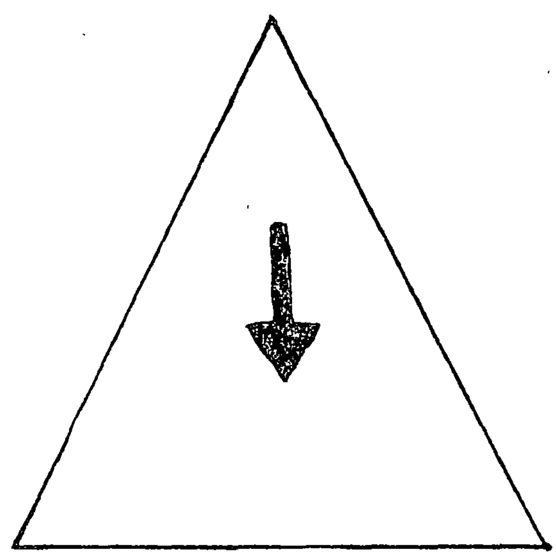
POLITICAS

ESTRATEGICAS

TACTICAS

ADMINISTRATIVAS

OPERATIVAS



para determinar el impacto de una decisión a nivel nacional en el ámbito regional.

Los sistemas geográficos pueden obviar este problema, ya que se pueden tener sistemas de cubrimiento regional o nacional

Las clases de información donde la información requiere de la identificación geográfica son muchas y muy variadas. A efecto ilustrativo mencionaremos las siguientes:

Recursos Físicos	- Todo tipo de recursos naturales - bosques - lagos - tierras de cultivo
Recursos humanos	- población - médicos - maestros
Obras de Infraestructura	- puentes - líneas telefónicas - caminos - distritos de riego - pozos
Producción	- agrícola - ganadera

- forestal
- de servicios

En fin, se puede concluir que toda actividad humana se desarrolla dentro de un contexto geográfico, si bien hay algunos en los cuales este contexto no es significativo.

ETAPAS DEL GEOPROCESAMIENTO

Ahora bien, dentro del proceso de utilización de información geográfica tenemos tres etapas bien definidas que son:

ETAPA I.- Analizar las relaciones físicas entre los factores del medio ambiente.

- uso del suelo
- características físicas
- obras de infraestructura

ETAPA II.- Se complementa con información estadística y socioeconómica de bases de datos institucionales.

- población
- valor de la propiedad

ETAPA III.- Se analiza el efecto de las políticas y cambios propuestas en:

- el uso del suelo
- facilidades
- servicios

Todo con objeto de coadyuvar a la toma de decisiones a diferentes - niveles con respecto a la

- asignación de recursos
- ubicación de facilidades
- determinación del uso potencial del suelo
- análisis de la demanda de servicios

DESARROLLO DE UN PROYECTO

De lo anteriormente expuesto pueden verse algunas de las implicaciones del geoprocesamiento que harán que para que éste tenga éxito, se cumplan las premisas que se mencionarán posteriormente.

Las implicaciones son:

a).- el geoprocesamiento cruza límites funcionales y jurisdiccionales.

b).- la información puede provenir de una gran variedad de fuentes.

c).- los usos pueden ser multiples.

d).- puede requerirse de más información de la actualmente disponible.

e).- se requiere de asignación de recursos financieros, materiales y humanos.

De esta resultan los siguientes requisitos para la aplicación exitosa del geoprocesamiento:

a).- se debe formar un grupo representante de las áreas usuarias que determine tanto las necesidades que debe cumplir el sistema como la información que el sistema puede absorver directamente.

b).- apoyo a un alto nivel directivo que coordine a los usuarios y que puede decidir sobre la asignación de recursos al proyecto.

TIPOS DE INFORMACION.-

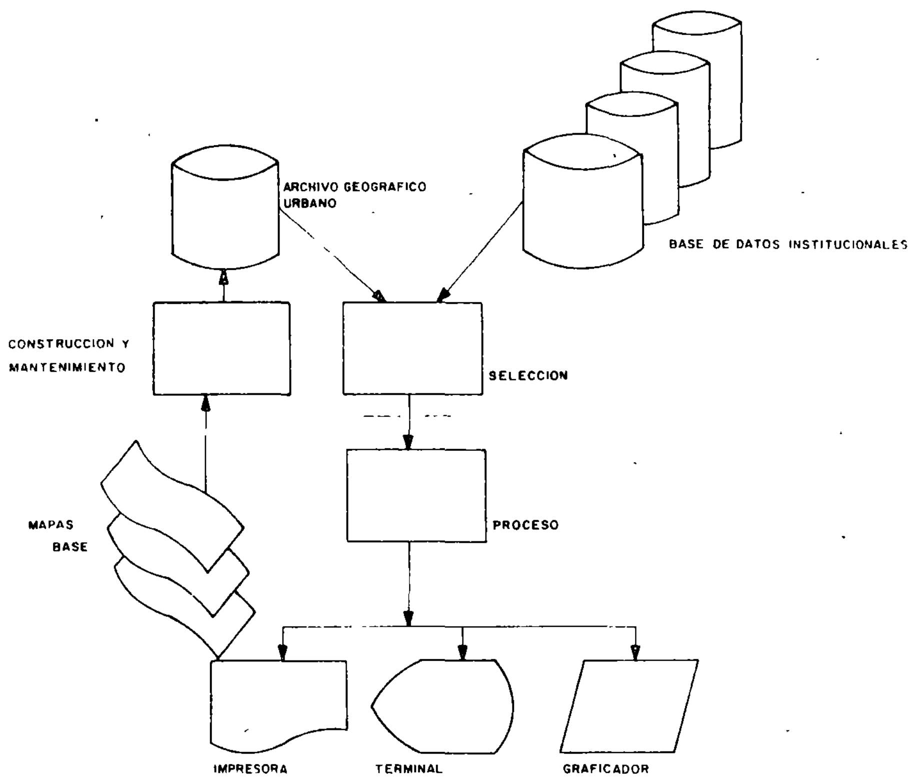
Básicamente, la información que puede utilizar un sistema de geoprocesamiento puede clasificarse como información:

Nominal - que es aquella que está localizada por nombres geográficos y cuyo ejemplo más relevante es el de la información censal.

Ordinal - que es la información localizada por coordenadas, como la representada en mapas.

De aquí resulta que los procesos del sistema son:

- construcción y mantenimiento de la base de


ARCHIVO GEOGRAFICO URBANO
BASE DE DATOS INSTITUCIONALES
CONSTRUCCION Y MANTENIMIENTO
SELECCION
MAPAS BASE
PROCESO
IMPRESORA
TERMINAL
GRAFICADOR

datos geográficos

- extracción de información, según el problema a resolver y proceso de la información.

Los resultados pueden darse en forma de listados de impresora o mapas realizados por impresora, resultados por pantalla, de caracteres alfanuméricos o de vectores y graficadores.

NIVELES DE DETALLE

De los diferentes niveles de detalle de la información se puede establecer una clasificación de sistemas de geoprocesamiento.

Los sistemas pueden ser a nivel regional o urbano dentro de cada uno se pueden tener los niveles de ingenieria, de límites, de localización o el geoestadístico.

Para entender mejor esta última división hablaremos de los diferentes niveles:

NIVEL DE INGENIERIA URBANO.-

Este nivel esta constituído por la información acerca de la posición de obras de infraestructura para proporcionar los servicios urbanos como por ejemplo, ubicación de líneas eléctricas, de agua potable y de drenaje dentro de una ciudad.

Generalmente las compañías de servicios públicos cuentan con mapas que indican la ubicación de sus facilidades. El objeto de geocodificar esta información puede ser por ejemplo, el de correlacionar la infor-

	INGENIERIA	LIMITES	LOCALIZACION	ESTADISTICO
NIVEL REGIONAL				
NIVEL URBANO				

mación de demanda de servicio con la de facilidades, con objeto de planear mejor el mantenimiento y extensión de las facilidades en función del crecimiento de la demanda.

NIVEL DE LIMITES URBANOS.-

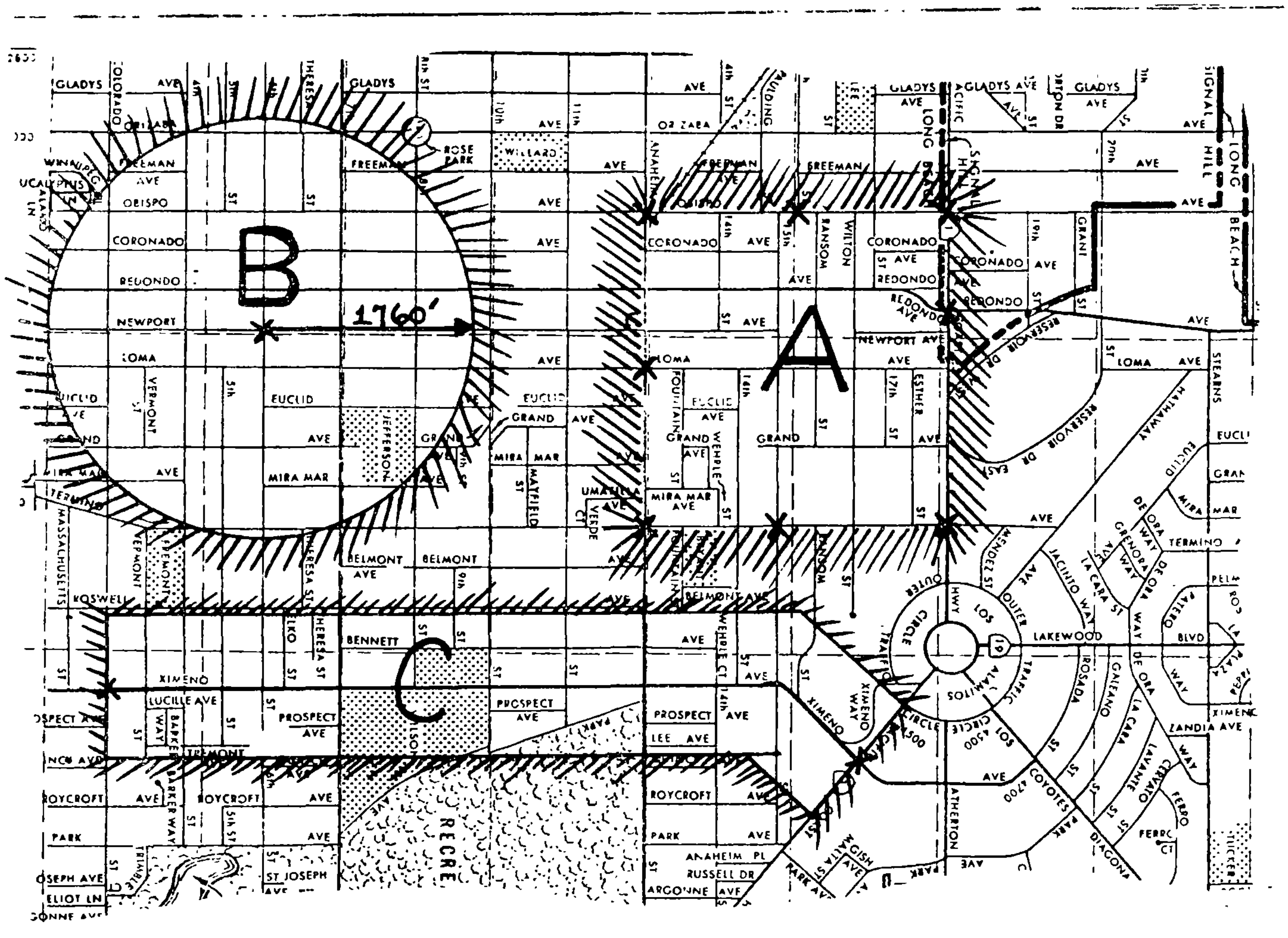
Este nivel es el resultado de un catastro urbano y el objeto de geocodificarlo puede ser por ejemplo, la determinación de los predios, y el valor de los mismos, que se verán afectados por la construcción de una nueva avenida. También puede servir para determinar a cuales -- predios se les puede revaluar en función de el mejor servicio que se les proporciona con esa avenida, etc.

NIVEL DE LOCALIZACION-URBANO.-

Es un nivel donde la información está agregada a nivel de manzana y puede servir para determinar el mercado potencial de un nuevo con-- junto de comercios o de agencias gubernamentales. Puede servir para efectuar una regionalización para efectos censales o de padrón de - votantes, etc.

NIVEL GEOESTADISTICO.-

En este nivel ya se tiene la información agregada a nivel de sección o cuartel y su uso puede enfocarse a la comparación relativa de diversas secciones en cuanto a nivel de población y de facilidades y servicios, etc.



A
B
C
1760'
GLADYS AVE
FREEMAN
OBISPO
CORONADO
REDONDO
NEWPORT
LOMA
EUCLID
GRAND
MIRA MAR
TERMINO
BELMONT
ROSWELL
BENNETT
XIMENO
PROSPECT
ROYCROFT
PARK
ST JOSEPH
ROSE PARK
WILLARD
JEFFERSON
ANAHEIM
RANSOM
WILTON
ESTHER
SIGNAL HILL
LONG BEACH
RESERVOIR DR
LOS ALAMITOS
TRAFFIC CIRCLE
LAKEWOOD BLVD
STEARNS
HATHAWAY
ATHERTON
COYOTES
RECRE

A nivel regional se pueden tener los mismos niveles de detalle de información: el nivel de ingeniería, el de límites, de localización y el estadístico.

NIVEL DE INGENIERIA RURAL.-

Este nivel puede considerarse constituído por modelos digitales del terreno. Se pueden tener MDT que cubren áreas muy extensas con poca precisión, lo que haría que no se pudiesen clasificar en esta categoría.

En algunos países se tienen bases de datos cartográficos a escala 1:2000, pero son realmente escasas.

NIVEL LIMITE-REGIONAL.-

En este sistema el nivel de información es el proporcionado por un catastro rural, que puede efectuarse a una escala entre 1:10 000 y -1:20 000. Aquí se tiene la delimitación de todos los predios rurales y su utilización puede ser semejante al correspondiente urbano, se puede utilizar para la determinación de los predios afectados por una presa o una carretera. Eventualmente, a través de la información recopilada a este nivel se podrían tener mapas de uso del suelo.

NIVEL LOCALIZACION-REGIONAL.-

En la actualidad, se están desarrollando bancos de datos de recursos que están a la escala 1:25 000 a 1: 250 000 elaborados a través de -

fotogrametría y fotointerpretación o por medio de clasificación automática de satélites.

Este tipo de Banco de Datos pueden servir para la planeación del uso del suelo, la determinación del área cultivada, etc.

NIVEL GEOESTADISTICO-REGIONAL.-

En este nivel, se empieza a usar una división política para agregar la información, tanto de recursos como socioeconómico. Las actividades - para las que puede usarse este tipo de Banco de Datos son genéricas. Es decir, podemos seleccionar entidades que cumplan una serie de requisitos, por ejemplo, todos aquellos cuya población sea mayor de -- 100000 habitantes y que sólo se tengan 10 hospitales, etc.

En general, el costo del desarrollo de los sistemas varía geometricamente según el nivel en el que se trabaje. Así, si un sistema geoestadístico para todo el país cuesta 1, el nivel siguiente costará 2, el siguiente 4 y el siguiente 8.

En general los geostadísticos manejan solo información nominal y los otros manejan información ordinal con cada vez mayor grado de detalle.

Las ventajas de los niveles más detallados es que siempre podemos - agregar información para alimentar el siguiente nivel, mientras que lo contrario es imposible.

Así, si se desarrolló un sistema de límites rural, con cost de 4 - podremos tener también sistemas de localización y estadísticos, que si se hubieren realizado independientemente nos costarían 3 unidades sin disponer del nivel de límites rural.

Es por lo tanto muy importante para un país determinar la convenien cia de empezar un sistema a nivel límites-rural o uno de localización, decisión nada fácil de tomar y que hasta ahora no se había planteado dado que el desarrollo de los sistemas geográficos de información que permiten la generalización automática se han desarrollado muy reciente mente y no hay conciencia clara de esta alternativa ni se han estudiado cuidadosamente sus factores e implicaciones.

CLASIFICACION

Los sistemas también pueden clasificarse en subsistemas de construcción y mantenimiento, de extracción y proceso.

A continuación se describen algunos sistemas típicos de algunos niveles:

a).- Banco de Datos Geográfico de DETENAL

La información que contiene este banco es la de los mapas elaborados por DETENAL sobre topografía, geología, edafología, uso del suelo y uso potencial.

El formato de la información almacenada permite tener una precisión variable de acuerdo a las necesidades de detalle de la zona geográfica - correspondiente. La actualización de estas propiedades se facilita por el diseño del formato y el usuario puede elaborar cualquier tipo de --

pregunta que se puedan descomponer en una combinación booleana de preguntas o predicados elementales. Su nivel es de localización regional.

Como ejemplo de las preguntas que el sistema puede responder tenemos:

¿ qué pueblos tienen más de mil habitantes y no tienen medios de comunicación?.

¿ cuáles son los mejores lugares para pasar una carretera, ¿ qué pueblos estan a menos de diez kilométros de una aeropista, ¿ cuántos kilométros cuadrados de pastizal inducido hay sobre terrenos de aluvión en una cierta zona?, etc.

La República Mexicana se dividió en 2400 zonas cada una de 1000 km^2., cada zona se dividió en 25 cuadros y éstos a su vez se dividieron en 12 subcuadros de 1' x 1', equivalente a 1.8 km., aproximadamente.

Para cada cuadro y de cada materia se codifican los siguientes tipos de propiedades:

Superficiales.- Son todas aquellas definidas por un área, como el tipo del suelo, la clasificación geológica o el uso del suelo.

Lineales.- Definidas por una línea que puede ser recta, curva o irregular como carreteras, líneas de telefonos, etc.

Particulares.- Son aquellas definidas por un punto, como por ejemplo, pueblos, depositos de agua, minas, etc.

Para la consulta se tienen funciones relacionadas y rutinas, además de funciones lógicas. Algunas de las funciones relacionadas son:

MAYORA.- Determina zonas, cuadros o subcuadros donde el porcentaje de una propiedad sea mayor que una cierta entidad.

MENORA.- Que se amenor

ENTRE.- Que el valor de la propiedad se encuentre entre 2 valores que se indican a continuación.

DIFERE.- Que el valor sea diferente.

IGUAL.- Que el valor sea igual.

BUSCA.- La función BUSCA nos indica la función de búsqueda y el nivel el cual se hace la consulta.

Algunas de las funciones lógicas son:

PRO (A1,A2,A3).- Búsqueda de propiedades.

donde:

A1.- Número de la propiedad buscada.

A2.- Función relacional.

A3.- Valor dado.

EVALUA.- Sirve para hacer una evaluación de las propiedades de un lugar.

CERCA (A1, A2,A3,A4).- Permite relacionar propiedades de diferentes - lugares y consta de 4 argumentos

A1.- Primer arreglo.

A2.- Segundo arreglo.

A3.- Distancia deseada.

A4.- Nivel de búsqueda.

PUEBLO.- (A1) Sirve para encontrar los pueblos por medio de un número previamente asignado.

HAYVIA (A1,A2).- Determina que tipo de caminos o comunicaciones que llegan a un publa.

A1.- Es el tipo de comunicación.

A2.- Es el número del pueblo.

SERPOB.- (NUM, a, b, c, d, e, f, g) Indica los servicios con que cuenta una población, de acuerdo a la clasificación de la carta de Uso del Suelo de DETENAL.

Todas estas funciones se pueden combinar por medio de parámetros - booleanos como el AND, OR, y NOT. Pudiendose construir preguntas - complejas.

b).- <u>Sistema Geomunicipal de Información.-</u>

Este sistema es del tipo geoestadístico-regional y la unidad básica de información es el municipio, por lo tanto, cualquier información que se desee alimentar a la base debe cumplir con la restricción de estar levantada a nivel municipal (1970).

Por medio de este sistema se pueden conocer la información de municipios o estados, construir conjuntos de municipios que cumplan con ciertas condiciones especificadas por el usuario, obtener reportes - con información y formatos seleccionados, crear formulas e índices y aplicarlas a los municipios. También se puede analizar los munici-- pios con respecto a sus vecinos inmediatos.

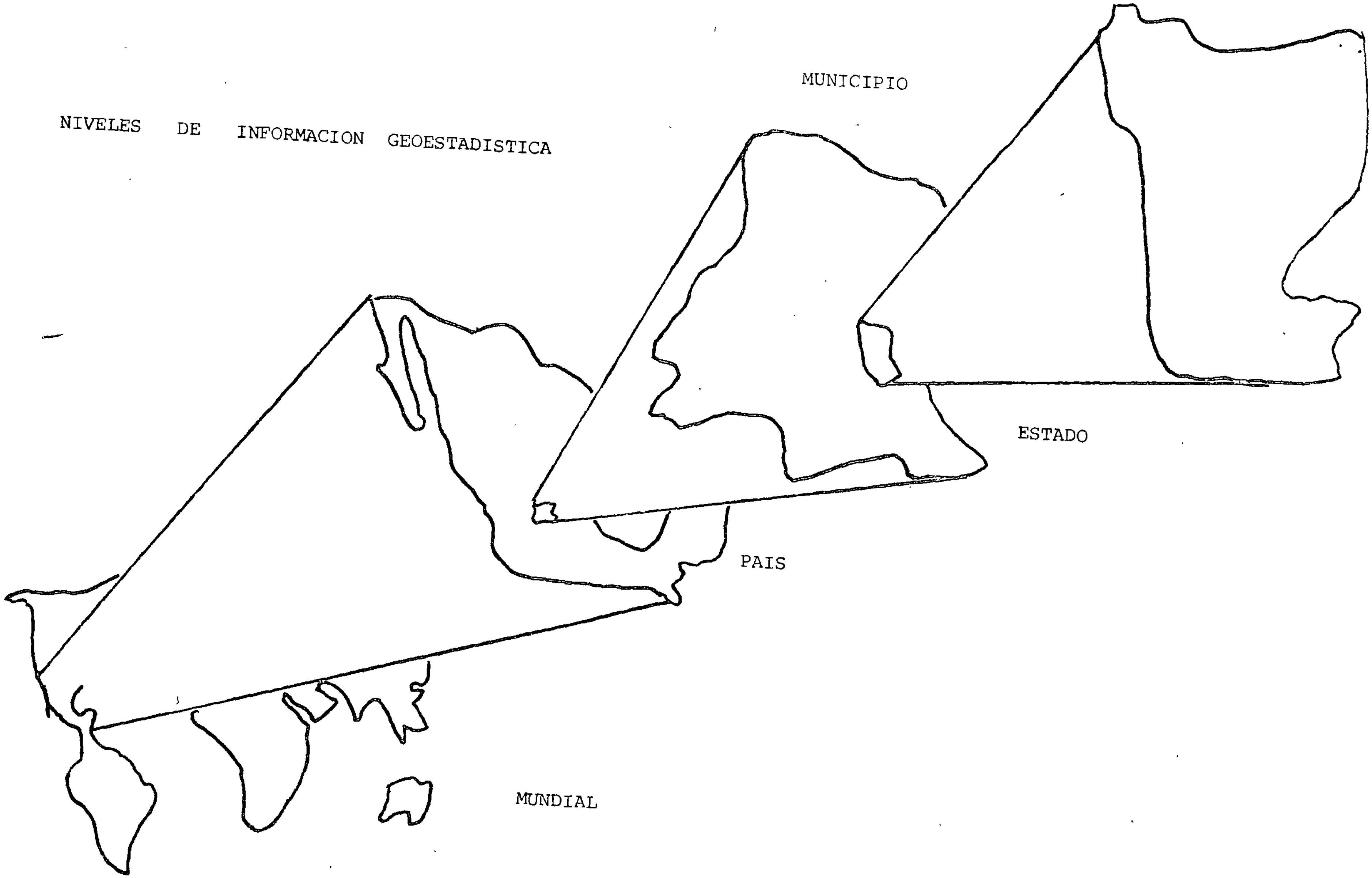
El sistema tiene las siguientes funciones:

CONSULTA.- Se tacha el número del municipio y a continuación aparecen los datos del mismo.

También se puede consultar por Estado.

VECINOS.- Nos da el nombre y clave de municipios vecinos.

NIVELES DE INFORMACION GEOESTADISTICA



CONJUNTO 1.- Agrega municipios que cumplan con una condición dada.

MUESTRA.- Nos da las claves de los municipios incluidos en un conjunto.

CONJUNTO 2.- Sirve para unir, intersectar u obtener el complemento de dos conjuntos existentes.

CONJUNTO MANUAL.- Permite definir un conjunto arbitrario.

DEFINE.- Permite construir expresiones algebraicas utilizando cualquier tipo de índice o variable definida anteriormente.

DESPLIEGA.- Muestra los conjuntos, entidades, formatos, campos o índices que están en los directorios.

FORMATO.- Para definir el formato de impresión y las variables que se imprimirán.

IMPRIME.- Imprimir un conjunto con un formato definido.

CARACTERIZA.- Se puede obtener para una variable o índice el valor total, máximo, mínimo y promedio.

ANALIZA.- Sirve para obtener un diagrama que indica la situación del municipio en relación a sus servicios con - respecto a un o más variables.

ESTRATIFICA.- Separa en diversos estractos un conjunto utilizando algunas variables o índice como criterio de estratificación.

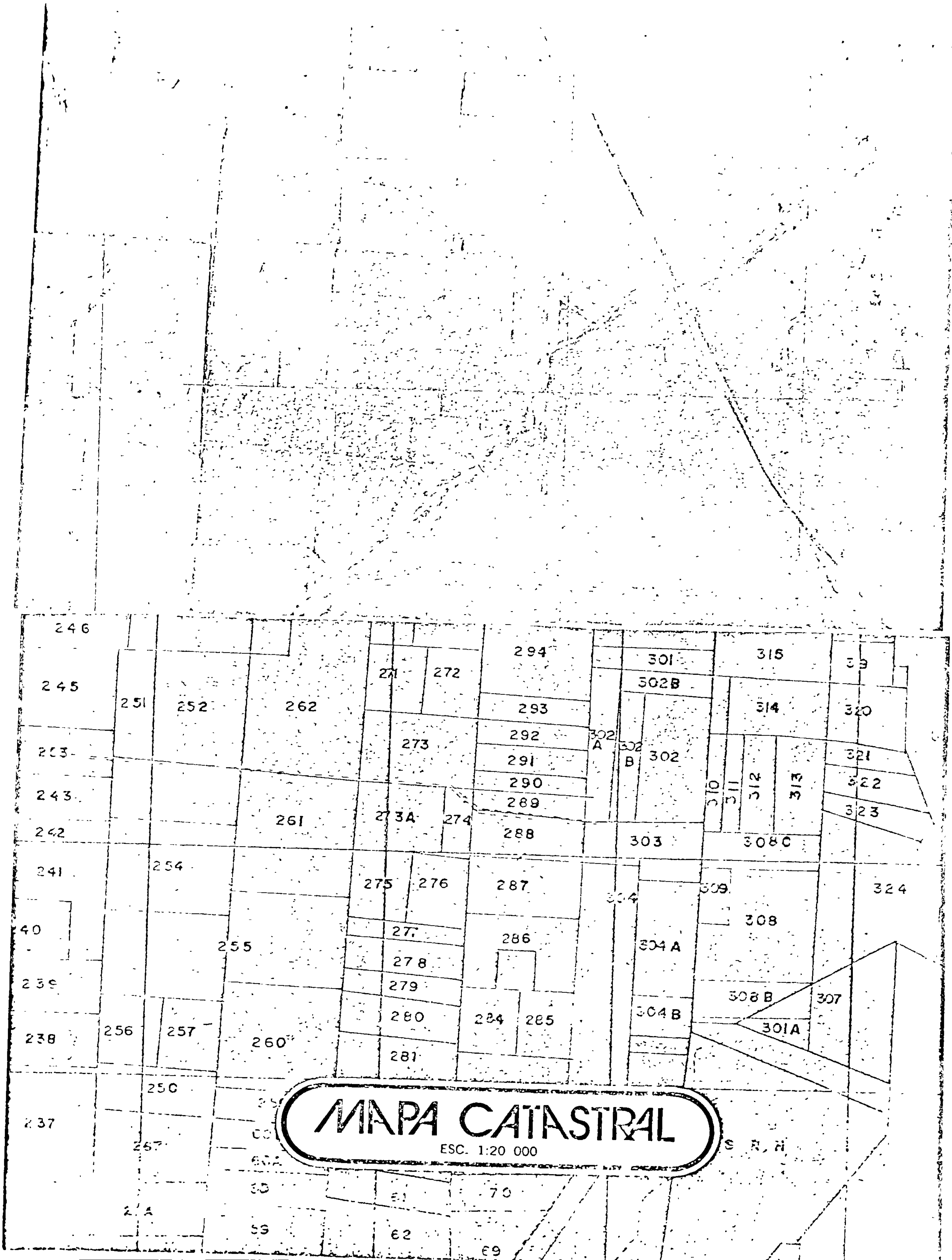
El sistema actualmente (el Sistema Geomunicipal) tiene información de los censos de población, agrícolas y ganaderos y de planeación educativa. Se le ha estado agregando otro tipo de información como por ejemplo la del Banco de Crédito Rural, etc.

<u>Sistema de Información de uso del suelo y registro de la propiedad de las provincias marítimas de Canadá</u>.-

Las provincias marítimas de Canadá se unieron para lograr los siguientes objetivos:

a).- Realizar la geodesia y mapeo integrado como base para - toda la información requerida para todo el desarrollo ligado al suelo.

b).- Reemplazar sus sistemas de registros manuales de la propiedad con un sistema en línea.


MAPA CATASTRAL
ESC. 1:20 000
246
245
251
252
262
271
272
294
301
315
302B
293
314
320
292
302 A
302 B
302
273
291
290
289
310
311
312
313
321
322
323
243
261
273A
274
288
303
308C
242
241
254
275
276
287
324
309
308
286
304 A
255
278
279
308 B
307
280
284
285
304 B
301A
256
257
260
238
281
250
237
70
62
69

c).- Examinar la viabilidad de un Banco de Datos del uso del suelo.

Para lograr lo anterior LRIS (Land Registration and Information - Service) se estableció en 4 etapas :

ETAPA 1.- Establecer los puntos de centros secundarios con mayor - densificación que le proporcionada por la Red Geodésica Nacional.

ETAPA 2.- Producir los mapas planimétricos, topográficos y de propiedades a las escalas requeridas para cumplir con las demandas generales por la administración de recursos, desarrollo urbano e identificación de propiedades.

ETAPA 3.- Reemplazar el sistema de registro de la propiedad con un nuevo sistema que aprovecha la tecnología moderna y la información de las fases anteriores.

ETAPA 4.- Establecer un banco de datos para las estadísticas geográficas, usando la información de los predios, obtenidas en las fases anteriores.

SISTEMA DE INFORMACION GEOGRAFICA URBANA

VISION GENERAL

Ing. Alberto Torfer Martell

Ana Elena Ferrer Ramírez

RESUMEN

Se presenta un sistema de información urbana con geocodificación que permite la integración de información de diversas fuentes, facilitando su agregación y presentación, especialmente en función de la distribución geográfica. Realizar este tipo de operaciones por medios manuales o sin un marco de referencia adecuado, consumiría cientos de horas hombre.

El sistema puede ser una herramienta muy útil tanto en la toma de decisiones administrativas como en la planeación.

C O N T E N I D O

CAPITULO I.-

INTRODUCCION

El medio ambiente ecológico, social y económico actual, está planteando demandas nuevas e imprecedentes, en las posibilidades administrativas y de planeación.

La explosión demográfica está provocando esfuerzos en nuestro medio ambiente. Durante los últimos 50 años se han producido más bienes, que durante todos los años anteriores registrados en la historia. Los vehículos están saturando las vías de comunicación. El aumento de fraccionamientos está inundando los sistemas de alcantarillado. El gran consumo de energéticos está contaminando el medio ambiente.

La efectividad de los gobiernos locales, se deteriora constantemente mientras las demandas siguen en aumento. Por ejemplo, la cantidad de basura por persona está aumentando constantemente. Al mismo tiempo, el tamaño y densidad del área metropolitana aumenta el problema de los basureros. En algunos casos se debe transportar la basura largas distancias para llegar a las áreas apropiadas. A esto se une el aumento de crímenes, necesidades de más escuelas, agua potable, alcantarillado, etc.

Todos estos problemas recaen en los administradores. Tienen que resolver los problemas duales del deterioramiento de la productividad y el incremento de demandas de nuevos servicios.- todo esto con presupuestos reducidos.

De acuerdo con estos problemas, se plantean interrogantes cada vez más complejas. Muchas de estas preguntas requieren datos de múltiples áreas geográficas y funcionales. Por ejemplo:

> ¿Qué impacto tendrá un nuevo fraccionamiento, sobre las escuelas, transportes, protección policiaca, drenaje, agua y alcantarillado?.
>
> ¿Cuál sería el impacto probable de una orden de reordenamiento urbano?.
>
> ¿Cómo se puede incrementar la productividad de los sistemas de protección contra incendios, con objeto de satisfacer las nuevas demandas?.

> ¿Cómo se deben desplegar los recursos actuales de patrullas policiacas para cubrir los cambiantes patrones sociales y criminales, sin sacrificar la seguridad pública?

Estos complejos problemas a nivel político no se pueden decidir con sistemas de información que únicamente proporcionan una orientación fucional. Se deben recolectar datos a través de fronteras funcionales y geográficas. Si no se tienen las respuestas adecuadas, es imposible la administración y planeación del medio ambiente ecológio, social y económico, actualmente complejo.

CAPITULO II.-

GEOCODIFICACION URBANA

La Geocodificación Urbana es una técnica reciente de procesamiento de datos, que se ha desarrollado para satisfacer los requerimientos de información de las agencias gubernamentales estatales y locales. Básicamente, relaciona la localización geográfica con las bases de datos existentes (por ejemplo, la localización geográfica se puede relacionar con los archivos de incidencias de policía, archivos de población escolar, etc.); la geografía es el común denominador que puede ligar virtualmente todos los datos. Los archivos del departamento de policía contienen información acerca de dónde y cuándo ocurren los casos. Los archivos de propiedad, así como los de permisos de construcción, contienen datos acerca de dónde se localizan las cosas. Los archivos de salud contienen información acerca de - quién y dónde, etc., y se podría continuar la lista indefinidamente.

El problema es cómo relacionar estos datos salvando el obstáculo que representa la diversidad de límites administrativos. Todas las agencias tienen sus propios métodos particulares para registrar su información. La solución sería relacionar de alguna manera esta información, con un común denominador (esto es, una localización geográfica). De esta manera, se podría recuperar toda la información perteneciente a esa localización, sin importar en qué base de datos se tenga. Esto es, en esencia, la geocodificación. La geocodificación urbana puede integrar y correlacionar los datos a través de los límites - administrativos de diversas dependencias, añadiendo de esta manera, toda una nueva dimensión al procesamiento de datos. Con la geocodificación, muchas dependencias pueden estudiar las bases de datos extremadamente diversificadas de otras, relacionando los datos en una base de datos geográficos común obteniendo respuestas a preguntas específicas. Anteriormente, bajo condiciones manuales, los complejos proyectos relacionados con el intercambio de datos entre diversas dependencias, hubiera sido prohibitivamente costoso en términos de hora-hombre; inclusive imposibles.

GEOCODIFICACION URBANA.-

La Geocodificación urbana puede ser vista como un proceso en tres pasos:

1) Utilizando mapas se analizan las relaciones físicas o geográficas entre las diferentes características relativamente estables del medio ambiente. ¿Cuáles son los usos básicos del suelo (por ejemplo, granjas, minería, residencial)? ¿Cuáles son las características físicas del área (Terreno, cuerpos de agua, tipo de suelo)? ¿Cuáles son algunos de los rasgos de infraestructura (presas, caminos)?

2) De nuevo, usando mapas, junto con datos estadísticos de varias bases de datos, se determinan los patrones de actividad generados por los diferentes usos del suelo. Cuáles son las rutas de tránsito colectivo, las poblaciones escolares, los valores de propiedad individual?

3) Se analizan los cambios propuestos para uso del suelo, facilidades o servicios que afectarán a estos patrones de actividad. Por ejemplo, ¿Cómo se pueden redistribuir las estaciones de bomberos para dar protección a nuevos edificios actualmente en construcción? ¿Cuál sería el efecto de un nuevo puente o carretera?.

Estos tres pasos podrán consumir incontables horas hombre si se hacen manualmente; se tendría que investigar en archivos enteros de diferentes departamentos. Con el geoprocesamiento de la información, una computadora podría hacer toda la búsqueda y correlación de los datos.

VENTAJAS DE LA GEOCODIFICACION

La ventaja más grande de la geocodificación es su potencial habilidad para incrementar la productividad; supera la suma de la utilidad de las aplicaciones individuales. Puede coadyuvar a mejorar la productividad operacional y de planeación de dependencias funcionalmente independientes, y a la eficiencia de todo el gobierno. A través de la mejor información, se podrá aumentar la eficiencia e inclusive, se podrán reasignar recursos. Esta posibilidad de reasignación es una de las ventajas principales del uso de estos sistemas.

Ya que la Geocodificación Urbana utiliza las bases de datos existentes, se podría perfeccionar la calidad total de las bases de datos individuales, de las siguientes maneras:

- Se podría eliminar la duplicación de una gran cantidad de datos, porque todos los identificadores geográficos serían mantenidos en la base de datos geográficos y no en los archivos de datos operacionales.

- Se podrían estandarizar los identificadores geográficos usados por todos los archivos de datos.

- Se podría checar la calidad e integridad de los datos almacenados.

- Se podría reducir el trabajo tedioso de codificación de identificadores particulares en las transacciones.

Cuando los funcionarios puedan evaluar exactamente las facilidades y servicios públicos requeridos por el desarrollo y puedan comparar estas demandas con los servicios y facilidades disponibles, entonces existirá una base válida para tomar decisiones. Se podrían consolidar fácilmente los datos de diferentes dependencias para formar una base general para toma de decisiones. Por ejemplo, la mayoría de los estados requieren de un informe del impacto del medio ambiente para un mejor desarrollo. Esto puede ser un requisito complejo que involucra datos de muchas dependencias independientes entre sí. ¿Cómo se afectará el medio ambiente? ¿Las facilidades existentes (escuelas, agua, etc.) serán capaces de soportar la carga adicional, o se tendrán que aumentar?. Las respuestas a estas preguntas requieren de la interrelación de una gran variedad de datos locales y ésta es otra ventaja del procesamiento geográfico - la de ser capaz de integrar datos a través de límites jurisdiccionales y funcionales.

La integración y correlación de la información a través de líneas funcionales, puede mejorar las siguientes áreas de política gubernamental y toma de decisiones:

- Asignación de recursos
- Ubicación de facilidades
- Uso del suelo
- Análisis de mercados
- Trabajos de ingeniería relacionados con los puntos anteriores.

Todas estas actividades son esenciales para la planeación de un mejor desarrollo. Todas las fases del desarrollo podrían ser cuidadosamente estudiadas por los planificadores, utilizando un sistema de procesamiento geográfico. Por tanto, se podrían trazar vastas redes de caminos, cañerías y líneas de energía para las necesidades óptimas presentes y futuras. Se podrían localizar eficientemente las estaciones de policía y bomberos, así como otras oficinas municipales. Se podrían localizar hidrantes para casos de incendio con objeto de dar la mejor protección posible a casas y oficinas. Se localizarían convenientemente escuelas y tiendas. Las líneas de transporte colectivo prodrían ser construidas donde más se necesitaran. Si se hiciera manualmente, cada proyecto representaría un enorme trabajo en términos de tiempo, dinero y recursos humanos. Sin embargo, un sistema de información urbano, podría correlacionar todos los datos necesarios para apoyar decisiones y políticas de importancia.

CAPITULO III.-

PRESENTACION GENERAL DEL SISTEMA

Un sistema de información urbana con geocodificación puede pensarse como una herramienta cuyo valor depende completamente de la habilidad y experiencia del usuario. El sistema capacita al usuario para agregar y desplegar rápidamente datos geográficos. Proporciona información para ayudar a los directivos y administradores a tomar decisiones efectivas. El sistema no toma las decisiones por sí mismo.

Los conceptos básicos de la geocodificación pueden ilustrarse mejor a través de un ejemplo: Evaluar el impacto de aplazar la construcción de un sistema de alcantarillado, en un área donde el crecimiento excede a los servicios; en otras palabras, las facilidades de alcantarillado - actuales no pueden satisfacer un crecimiento adicional. Un retraso - sería una de las posibles opciones para enfrentarse al problema, hasta que se pudieran construir facilidades adicionales que abarataran el costo total del proyecto.

La geografía del área se representa como una serie de mapas sobrepuestos.

1.- El primero (Figura 1) muestra la geografía física relacionada de un área hipotética de planeación; el río, los arroyos, y sus cuencas de drenaje asociadas. Las líneas punteadas representan los límites de las cuencas. El significado de las cuencas de drenaje es que cada una está servida por una sola red de alcantarillado, con una capacidad - dada.

2.- La segunda sobrepuesta (Figura 2) muestra las facilidades físicas o de infraestructura. En este ejemplo, se muestra la red de carreteras, Otras sobrepuestas podrían mostrar escuelas, tiendas, áreas recreativas, etc.

3.- La tercera sobrepuesta (Figura 3) muestra los límites pólíticos del área. En este ejemplo, zonificación de distritos (comercial rodeada por residencias, rodeada por agricultura). Otras sobrepuestas podrían involucrar red de caminos, límites de parcelas, etc.

4.- La siguiente sobrepuesta (Figura 4) muestra los límites administrativos. En este ejemplo, se muestran las zonas escolares, pero también se podrían usar distritos de bomberos, distritos electorales, etc. En esta etapa, el mapa muestra la fisonomía física, facilidades de - infraestructura, límites políticos y límites administrativos. Con esta descripción básica de las relaciones físicas y geográficas, se podría hacer manualmente un análisis elemental.

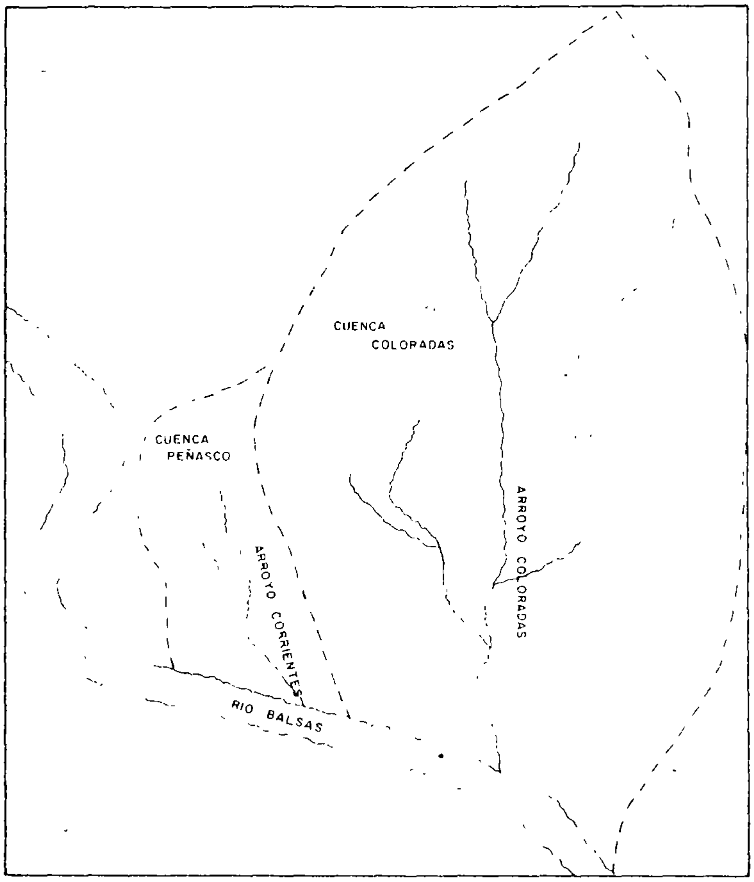


Figura 1 EJEMPLO DE UN MAPA TOPOGRAFICO SIMPLIFICADO

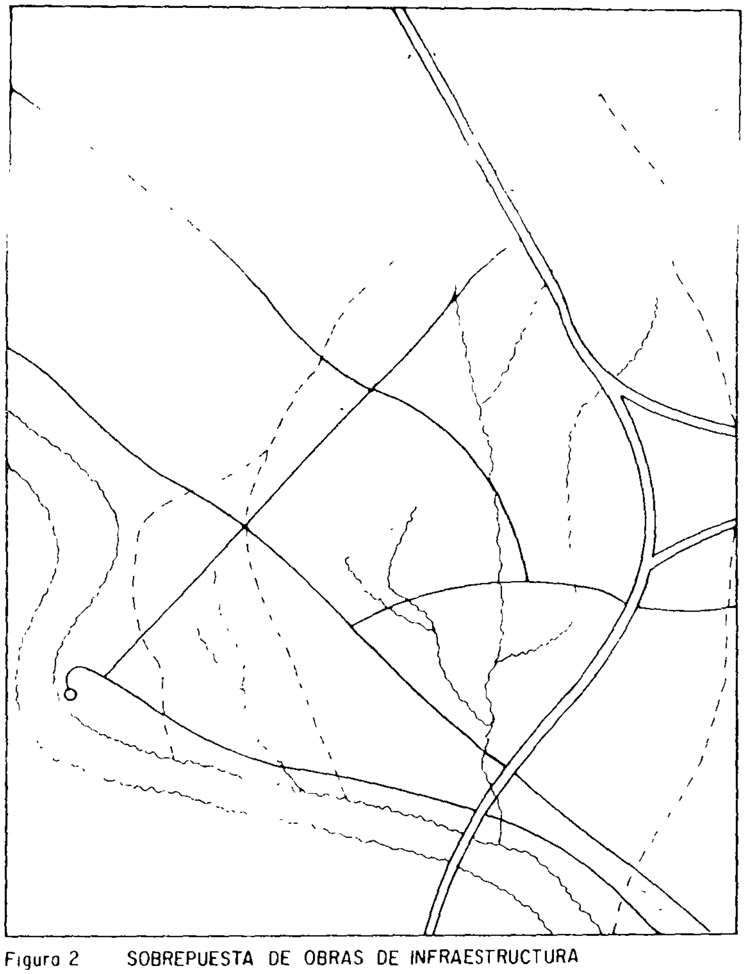


Figura 2 SOBREPUESTA DE OBRAS DE INFRAESTRUCTURA

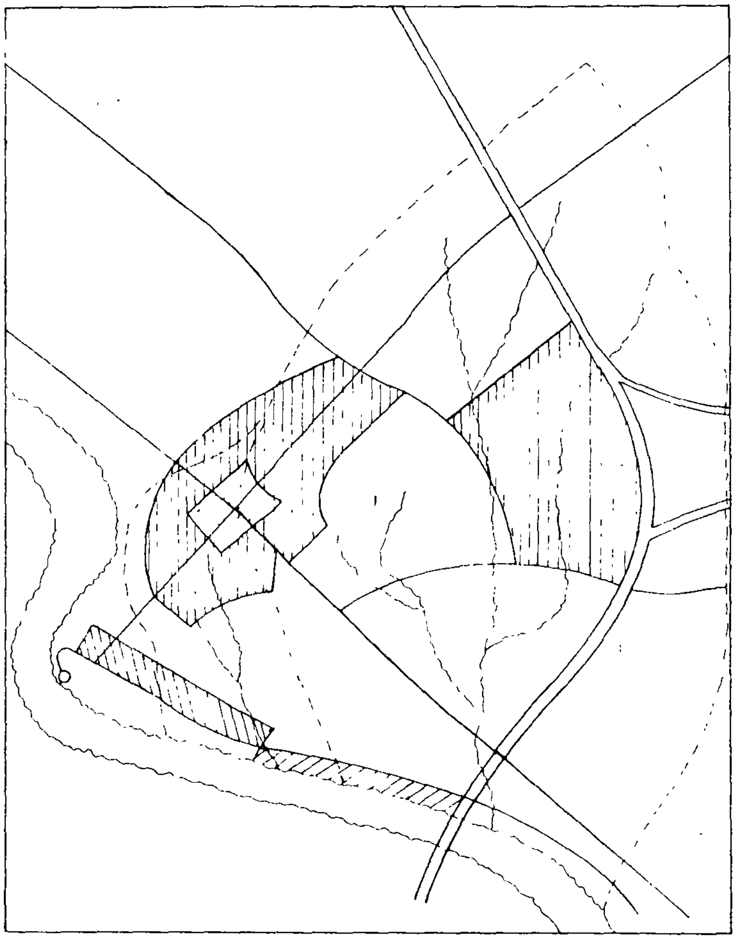


Figura 3 SOBREPUESTA DE LIMITES POLITICOS

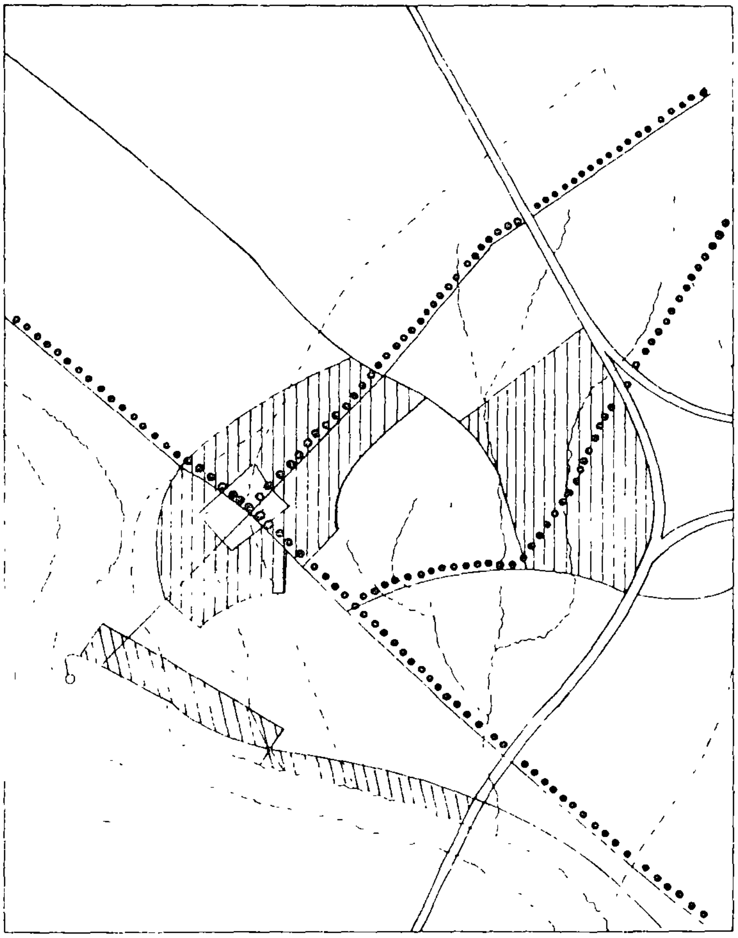


Figura 4 SOBREPUESTA DE LOS LIMITES ADMINISTRATIVOS

Llevando el problema un paso más adelante, esta geografía puede relacionarse con otros archivos de datos, como archivos de permisos de construcción, archivos de casos de incendio, policiacos, etc., para lograr un análisis más preciso de la situación. Para este problema en particular, se saca del archivo de predios, una lista de predios vacantes. Los predios vacantes se relacionan con las cuencas de drenaje y límites de zona y se hace un proyecto del potencial de construcción. El número y tipo de construcciones propuestas llevan a un proyecto preciso de la capacidad potencial de alcantarillado requerida para cada cuenca de drenaje. El impacto del retraso del alcantarillado se podrá calcular con precisión.

(Un análisis diferente podría ser el impacto sobre la inscripción escolar proyectada, utilizando límites de zonas escolares en lugar de los límites de cuencas de drenaje).

Manualmente esta aplicación, podría consumir varios cientos de horas-hombre. Muchas dependiencias hacen análisis de este tipo, únicamente para problemas de rutina, y no están relacionados con proyectos especiales. Este procedimiento completo, y muchos otros semejantes, pueden automatizarse utilizando un sistema de información geográfica urbana.

La figura 5 muestra un modelo simple de un sistema de procesamiento geográfico. Los mapas alimentados al sistema son los sobrepuestos ya mencionados. Las características físicas, obras de infraestructura, límites políticos y áreas administrativas. El procesador geográfico combina estas entradas para producir reportes o mapas que apoyen al proceso de toma de decisiones.

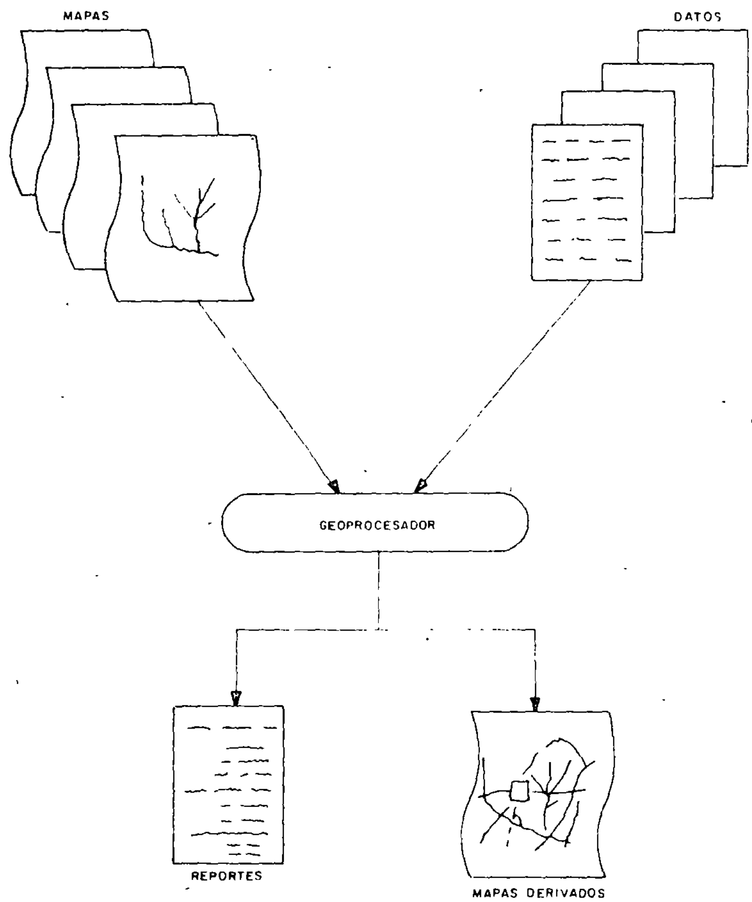


Figura 5 MODELO SIMPLE DE UN SISTEMA DE INFORMACION URBANA

CAPITULO IV.-

DESCRIPCION DEL SISTEMA

Un sistema de geoprocesamiento se puede dividir en tres fases distintas: Construcción/Mantenimiento, Selección y Proceso (ver figura 6).

CONSTRUCCION/MANTENIMIENTO.-

La construcción/mantenimiento, tiene la función de construir, editar y actualizar el archivo geográfico urbano(AGU). Este AGU es una descripción digitalizada de la geografía de un área específica, junto con sus índices asociados (incluyendo el número de calle y/o parcela). El AGU es un modelo de computación de un área geográfica, casi del mismo modo que un mapa es un modelo gráfico.

DIGITALIZAR un mapa es poner los datos en una forma legible de máquina, con lo cual se crea un archivo de procesamiento de datos del mapa. Esto se logra asignando coordenadas X, Y, a los puntos predeterminados en el mapa. Por ejemplo, se les pondrían coordenadas X, Y a las intersecciones de calles, así como a otros rasgos, como localizaciones de hidrantes para incendios, cementerios, puentes y carreteras. Entonces se le da un nombre a cada uno de estos puntos. Esta lista de nombres forma el índice de calles que se asocia a el AGU y se puede considerar como un "diccionario de datos". Cada nombre muestra la información - disponible para cada punto y dónde se puede encontrar esa información en la base de datos.

SELECCION, la Selección busca datos apropiados en el archivo geográfico urbano, y los correlaciona con los archivos de datos existentes (departamento de policía, consejo de planeación, asesoría, etc.) Por ejemplo, en el AGU se puede correlacionar un inventario de carreteras, un archivo de escuelas primarias, y una base de datos de parcelas, para analizar los sitios adecuados para construir una nueva escuela. La correlación se realiza haciendo coincidir los identificadores de localización que aparecen en el AGU y en las bases de datos operacionales.

PROCESO, el Proceso organiza los datos en el formato deseado y despliega el resultado. La producccción puede ser gráfica (mapas), alfanumérica (tabulación estadística o de terminal interactiva). Esta producción se puede hacer en impresoras, graficadores, terminales remotas, o terminales gráficas de vectores. El modelo se puede manejar en tiempo real.

Los mapas se pueden producir en una impresora de líneas, un graficador o un tubo de rayos catódicos (pantalla de video).

Para el público se pueden hacer mapas con graficadores de alta calidad, mientras que para otros mapas que no necesitan ser tan refinados se - puede usar impresoras electrostáticas.

El proceso puede contener un modelo para preguntas: de "qué pasaría si". Por ejemplo:

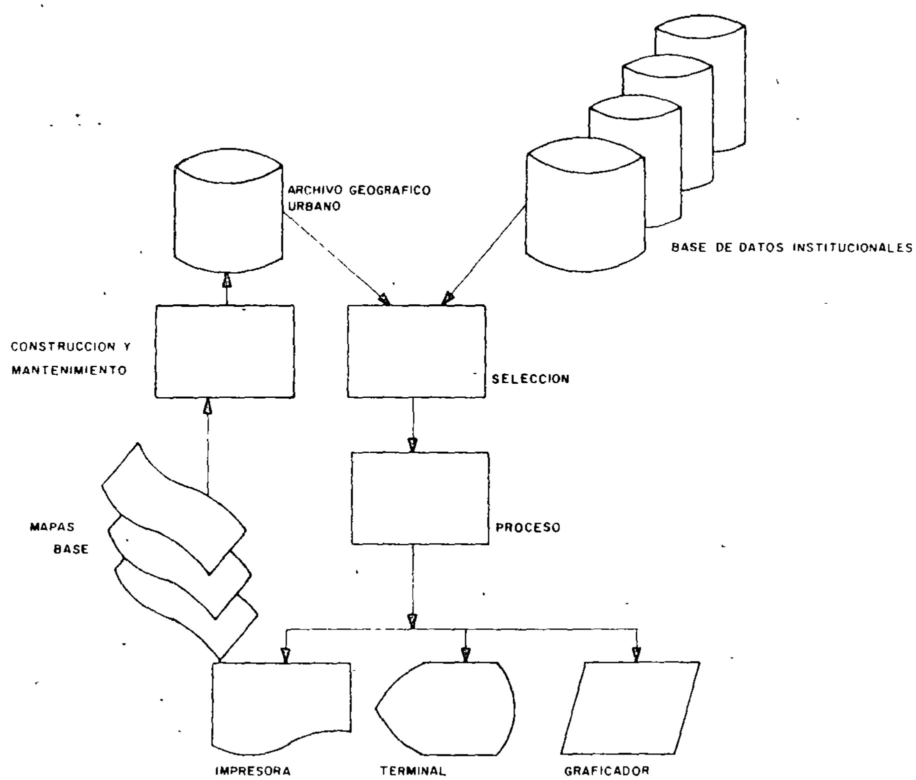


Figura 6 _ UN SISTEMA DE GEOPROCESAMIENTO

- El modelo de análisis de tráfico evalúa el nivel adicional de tráfico resultante de un nuevo conjunto comercial en el lugar. Se usa un modelo interactivo por medio de una terminal, donde se pueden fácilmente añadir, suprimir o cambiar datos. (¿Qué pasaría si se construyeran 13 tiendas en lugar de 8? ¿Cuál sería el impacto?).

- El modelo de contaminación del aire computa las concentraciones de contaminantes gaseosos (hidrocarbonos, monóxidos de carbono, etc.), que resultan de los equipos móviles y estacionarios que contaminan el ambiente. (¿Qué pasaría si hubiera 400 automóviles en lugar de 50? ¿Cuáles serían los niveles de contaminación?).

- El modelo de ordenamiento territorial mide el impacto de nuevos fraccionamientos sobre los servicios públicos, escuelas, facilidades recreativas, de salubridad y facilidades de proyección contra incendios, etc. (¿Qué pasaría si se constuyeran habitaciones para 15 familias adicionales? ¿Cuál sería el impacto?).

Estos tres modelos utilizan un archivo geográfico urbano que se correlacionan con las bases de datos para un sistema interactivo que posibilita a los administradores y planificadores resolver los problemas actuales y asegurar las necesidades futuras. Se puede añadir, cambiar o suprimir cualquiera de las partes de la información por medio de una terminal, y el programa se puede volver a correr para las determinaciones del último minuto.

CAPITULO V.-

MANEJO DE LA BASE DE DATOS

El Sistema de Información Geográfica Urbana correlaciona los datos de las bases institucionales, con el archivo geográfico urbano (figura 7).

Los archivos geográficos urbanos se pueden construir en cuatro niveles de complejidad; explícito, estadístico, de límites, y de ingeniería. A medida que se procede desde el básico hasta el más complejo, se pueden manejar una variedad mayor de aplicaciones; también se incrementa el costo y los requerimientos de mantenimiento del sistema.

El nivel de complejidad explícito, es el acceso básico al procesamiento geográfico (figura 8). En este nivel, cada registro de datos contiene el geocódigo. Por ejemplo, el distrito de bomberos está contenido en cada reporte de incidencia de incendios; o la cuenca de drenaje está contenida en cada registro de predio. Con este nivel, se puede hacer un procesamiento geográfico simple, sin el auxilio de un sistema de computación especial. Empero, existen los siguientes inconvenientes:

1.- No es posible cruzar la información de diferentes límites funcionales, a menos que todas las bases de datos estén geocodificadas.

2.- Es virtualmente imposible actualizar los archivos geográficos cuando se cambian los límites administrativos por una nueva zonificación, división, etc.

3.- Las preguntas de "qué pasaría si", no pueden ser económicas porque las fronteras geográficas propuestas no están en el archivo. Por estas razones, el nivel de complejidad explícito no se utiliza tan ampliamente en los sistemas de procesamiento geográfico.

ENFOQUE MULTIDISCIPLINARIO

A las tres distintas fases de un sistema determinado de información urbana (Construcción/Mantenimiento, Selección y Proceso) se les puede denominar el enfoque multidisciplinario. Existen tres niveles de complejidad, (en realidad existen cuatro, pero no se incluye el nivel explícito por su simplicidad) por localización (direcciones y geocodificación simple), por límites (administrativos, políticos o debidos a la malla urbana) y el de facilidades, que defieren entre sí, principalmente en el nivel de detalles geográficos en los archivos geográficos urbanos.

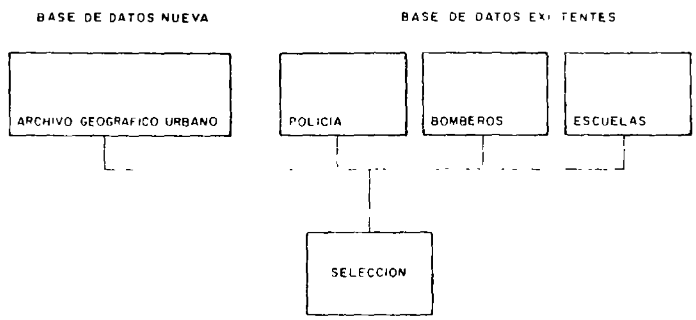


Figura 7 CORRELACION DE INFORMACION FUNCIONAL

NUMERO DE PREDIO	ZONA ESCOLAR	DELG POLICIA	DIST PLANEA	Z ALCANTA	P INUNDA	DATOS PREDIO
10 04 - 1 - 24 1	321	512	5	4	0	
10 05 - 1 - 24 2	321	512	5	4	1	
17 65 - 3 - 14 21	625	327	7	3	1	
17 65 - 3 - 14 22	626	327	7	3	1	

Figura 8 EJEMPLO DE UN METODO DE GEOPROCESAMIENTO

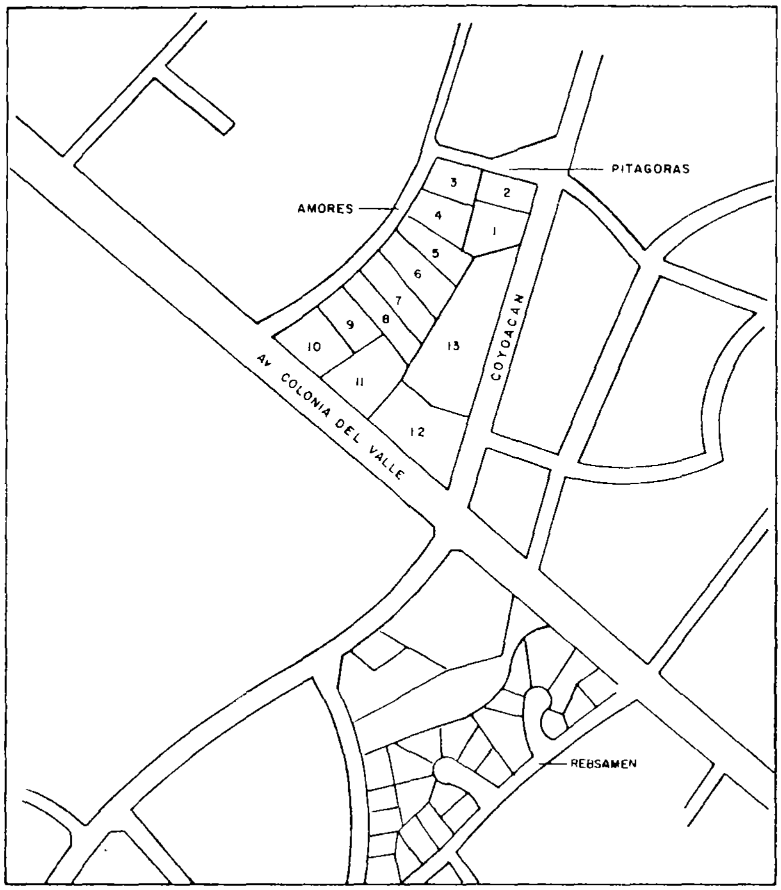


Figura 9 EJEMPLO DE UN MAPA PREDIAL

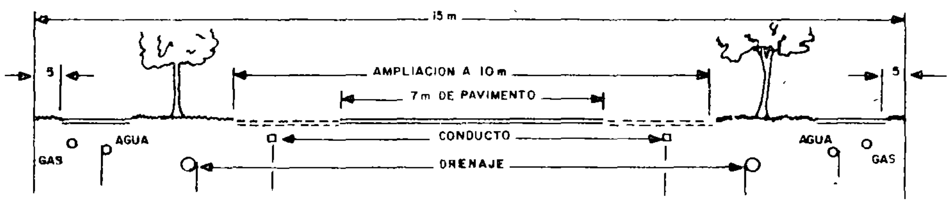


Figura 10 EJEMPLO DE UN DISEÑO DE INGENIERIA DETALLADA

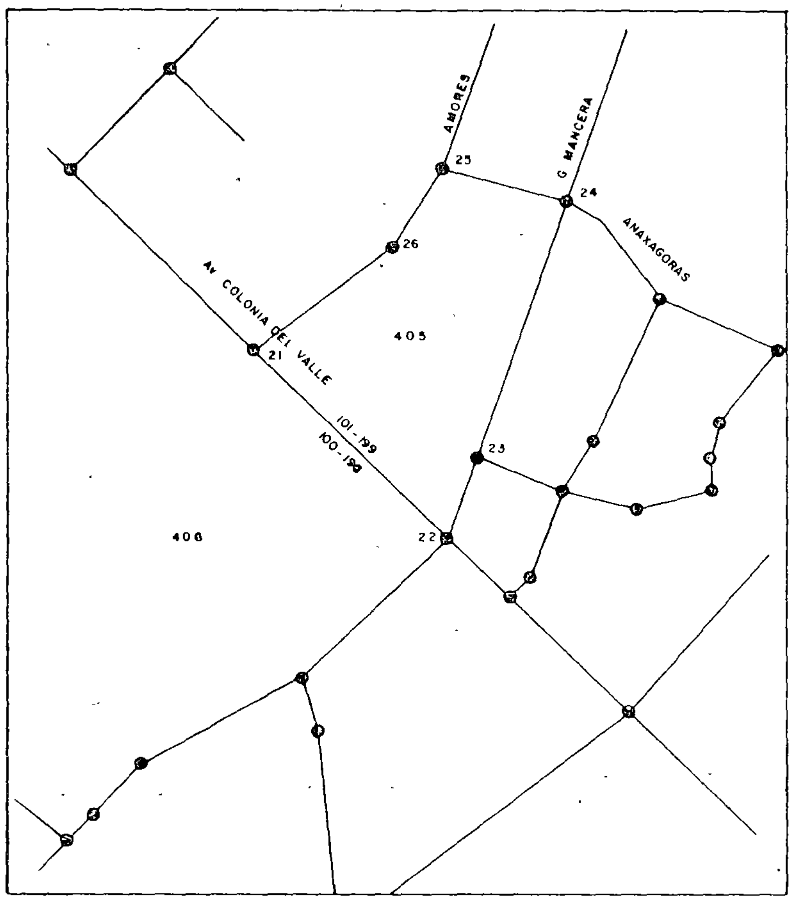


Figura II EJEMPLO DE UN MAPA AGU/CDMU

1.- El nivel de localización simple resuelve la ubicación de acontecimientos en áreas geográficas (por ejemplo, en el distrito de bomberos 14 ocurrió un incendio en Ave. Taxqueña 181, o en el área de planeación 5 ocurrió una subdivisión del predio 10-1-24). Responde a la cuestión básica: "¿En qué áreas geográficas ocurrieron los hechos?.
Tiene la posibilidad de producir reportes estadísticos que sirven para el análisis o respuestas interactivas para una acción inmediata. Este nivel también puede apoyar la producción de mapas de incidencia o de densidad.

2.- El de límites resuelve la geometría de los acontecimientos, en otras palabras, las intersecciones actuales de los polígonos. (Por ejemplo, en la figura 9. ¿Qué porción de qué predio cruzará exactamente el trazo propuesto para una nueva calle o carretera?.) El nivel estadístico listaría solamente los predios afectados. El sistema de límites da una respuesta geométrica más que un sumario estadístico. Este nivel sería útil para el valuador catastral; por ejemplo, podría pedirse contribución especial a todos los predios a 50 metros de una nueva avenida?. Se podrían preparar planos catastrales, para aquellos predios que estén incluídos, aunque sea parcialmente.

3.- El nivel de facilidades (o Ingeniería) resuelve los detalles a nivel de ingeniería (por ejemplo, porporciona trazos de ingeniería sin requerir de un levantamiento preliminar). Ver figura 10.

Todos estos niveles se pueden usar en el Enfoque Multidisciplinario, y son compatibles en nivel ascendente. Cada nivel más alto permite una mejor visión, pero es más complejo, y por consiguiente más costoso en su construcción y mantenimiento. Generalmente es más práctico empezar con el nivel de direcciones (localización) y más tarde extenderse hacia arriba de acuerdo con las demandas que se presenten. El Enfoque Multidisciplinario apoya las aplicaciones en tiempo real, como los despachos de emergencia, así como las aplicaciones de planeación.

ARCHIVO GEOGRAFICO URBANO/CODIFICACION DUAL INDEPENDIENTE PARA MAPAS URBANOS.-

Ya son varios los países que han implementado un sistema de información para áreas urbanas cuyas características principales son:

> Una retícula de canevá en proyección cónica de Lambert para el caso de Suecia, sirve de base a la incorporación de información en la base de datos (FRIS).
>
> En Canadá la Unidad Básica de Información está constituida por elementos lineales representado un costo del bloque de manzana (G.R.D.S.R.) y en el caso de los países escandinavos (NIMS) E.U.A. (DIME) y Francia (R.G.U.) la Unidad Básica es el segmento vial limitado por dos nodos o cruceros.

La Oficina de Levantamiento Censal de E.U.A. ha creado un archivo geográfico urbano para las áreas metropolitanas más grandes del país. Estos archivos se utilizarán para el Censo de 1980 y algunos del Censo de 1970, ya están disponibles. Ver figura 11. Sin embargo, algo más importante es que estos archivos (AGU/CDIMU) tienen los datos básicos para apoyar un sistema de procesamiento geográfico en el nivel de localización. La existencia de un archivo AGU/CDIMU para un área en particular, puede reducir bastante el tiempo y gastos necesarios para construir un sistema de procesamiento geográfico.

Los usuarios de estos sistemas serán incontables y ayudarán a enriquecer las bases de información de datos urbanos multidisciplinarios que existan.

CAPITULO VI.-

DESARROLLO DEL SISTEMA

El desarrollo de un sistema de información urbana es una empresa compleja. Es absolutamente esencial un apoyo a nivel directivo. Ya que el procesamiento geográfico cruza los límites funcionales, debe existir un apoyo a alto nivel, para cuando las decisiones crucen estos límites.

Se debe organizar un grupo con representantes de las dependencias que vayan a usar el sistema. Estos representantes participarían en la planeación de todo el sistema, con objeto de asegurar la satisfacción expedita de las necesidades y requerimientos de todas ellas. Se debe consultar a los usuarios potenciales durante el diseño del sistema, con el fin de conocer si aceptarán el producto terminado.

Este grupo de representantes de cada área debe llevar a cabo un estudio de lo siguiente:

- Aplicaciones específicas deseadas para cada dependencia (incluyendo los beneficios potenciales).

- Bases de datos disponibles y las que se tendrán que construir para apoyar las aplicaciones de cada unidad.

- La investigación y la transferencia de códigos o técnicas de potencial aplicación al sistema.

- Los recursos disponibles tanto locales, estatales o federales.

Entonces se procede a la elaboración de un plan de implementación basado en las aplicaciones requeridas por cada área y los recursos disponibles. Se podrían añadir al sistema las capacidades y aplicaciones, de manera modular, estableciendo un programa de implementación.

CAPITULO VII.-

JUSTIFICACION ECONOMICA

El valor real de un sistema de información urbana, se debe medir en términos del aumento de productividad en todas sus aplicaciones.

Se puede analizar la productividad desde tres puntos de vista:

1.- Reducción de costos.- Eliminación de algún desembolso de capital en el futuro. (Por ejemplo, evitar la construcción de una nueva estación de bomberos, porque el equipo actual y los recursos humanos se aprovechen más eficientemente).

2.- Redistribución de costos, recursos humanos, o equipo.- A través de una asignación más eficiente.

3.- Valor agregado.- Resultado de las posibilidades adicionales - que antes eran muy costosas o muy difíciles de realizar. La posibilidad de establecer una regionalización nueva cae en esta categoría; frecuentemente, los requerimientos federales y estatales demandan informes sobre el impacto en el medio ambiente para nuevos proyectos, y éstos pueden ser muy complejos. Las horas-hombre empleadas en la investigación manual, podría ser astronómicas.

APLICACIONES POTENCIALES DEL PROCESAMIENTO GEOGRAFICO

En toda administración municipal existen aplicaciones del procesamiento geográfico. Continuamente se hacen algunas aplicaciones (como despachos de emergencia de bomberos o policía). Otras, como la regionalización se hacen sólo ocasionalmente como una función de la planeación. Todavía - otras, como las inspecciones y permisos de construcción, se hacen rutinariamente. A continuación se presenta una lista de las aplicaciones potenciales del sistema de información urbana para poblaciones de - 100,000 habitantes en adelante. No pretende ser completa, únicamente indicativa.

1.- Despacho de policía.- Despacho de recursos policiacos basados en la asignación geográfica y la historia local.

2.- Despacho de bomberos.- Despacho de recursos de protección contra incendios, basado en la localización geográfica, condiciones locales, y la historia del caso.

3.- Zonificación escolar.- Planeación de la utilización óptima de las facilidades educacionales, basada en la geografía de la población escolar.

4.- Transporte escolar.- Planeación de las rutas de autobuses más eficientes, basada en la población escolar y la localización.

5.- Planeación urbana.- Análisis del impacto de las decisiones urbanas sobre todas las dependencias del gobierno local.

6.- Planeación de transporte.- Planeación del transporte colectivo más eficiente, modificando las rutas y destinos de los transporres alimentadores.

7.- Recreación.- Localización de las mejores áreas recreativas, basada en la localización geográfica de los sitios y usuarios.

8.- Análisis de delincuencia.- Asistencia en la delimitación de áreas de alta incidencia de delitos para optimizar la protección policiaca.

9.- Cuidado de parques públicos.- Planeación del cuidado de zonas verdes dentro de la ciudad, basada en la localización de estas zonas.

10.- Colecta de basura.- Planeación de rutas, basada en la geografía del uso del suelo.

11.- Impacto en el medio ambiente.- Análisis del impacto de las urbanizaciones propuestas, sobre la ecología y las dependencias municipales.

12.- Listas de empadronamiento.- Producción de listas de electores por área geográfica.

13.- Localización de servicios públicos.- Planeación de reparaciones y nuevas construcciones de líneas eléctricas, telefónicas, de agua potable, alcantarillado, etc.

14.- Uso del Suelo.- Mantenimiento de archivos exactos de uso del suelo.

15.- Monitores del crecimiento.- Análisis contínuo de los cambios de localidad, basado en los análisis geográficos de licencias y permisos de construcción.

16.- Inspecciones de construcción.- Planeación de programas de inspección, basada en la distribución geográfica de permisos y licencias.

17.- Mapas base para Catastro Urbano.- Determinación de los impuestos prediales.

18.- Valuaciones especiales.- Asistencia en el trazo de áreas candidatas a impuestos especiales.

19.- Mapa base (Ingeniería).- Mantenimiento de mapas de ingeniería exactos para eliminar la necesidad de estudios preliminares.

No todas las aplicaciones se usarán en todas las organizaciones; de cualquier modo cada aplicación sirve para, por lo menos, dos dependencias. Cada aplicación tiene un cierto potencial de productividad. La suma total de la productividad justificará un sistema de información geográfica urbana como el que se plantea.

CAPITULO VIII.-

CONCLUSION

La implementación de un sistema de información geográfica urbana puede dar muy buenos resultados para los gobiernos estatales y locales. Como con la implementación de cualquier sistema computarizado sofisticado, se necesita un apoyo completo y una gran confianza por parte de los altos niveles gubernamentales. La geocodificación integra información a través de límites funcionales (dependencias) y jurisdiccionales (políticos); el procesamiento geográfico no puede justificarse en base a una sola dependencia. Por tanto, la construcción, mantenimiento y uso del sistema, requieren de la completa cooperación y compromiso de todos los funcionarios, además del coordinador del proyecto. Sólo entonces se pueden obtener los recursos necesarios para tener éxito, facilidades, dinero y personal. Se debe consultar al funcionario de nivel más alto durante todas las fases de la planeación e instalación de la base de datos del sistema geográfico urbano.

'pg

MANUAL DE UTILIZACION DEL BANCO DE DATOS CETENAL

ING. ERNESTO BRIBIESCA CORREA

PREFACIO

El objetivo principal del Banco de Datos CETENAL (BDC), es resolver en forma rápida y confiable, gran cantidad y variedad de problemas sobre el mejor aprovechamiento de los recursos del país, que requieren el manejo de información abundante y compleja.

Actualmente, con la ayuda del BDC, se han localizado cuencas lecheras en el área de Ojo Caliente, Zac.[1] y en las áreas de Celaya, Querétaro, Cortázar, Apaseo el Alto y en el Estado de Aguascalientes se han señalado rutas para carreteras, se han localizado lugares aptos para desarrollar ciertos cultivos específicos, y se han propuesto lugares apropiados para una aplicación dada, con base en soluciones anteriores. Sin embargo, la solución de estos problemas apenas indica el inicio de una serie de aplicaciones en planificación regional, estudios económicos y otros aspectos.

Este manual hace posible el acceso del usuario al BDC, con un lenguaje de fácil aprendizaje y comprensión, que no requiere conocimientos previos en ciencias de la computación.[2]

[1] Véase referencia 2 pag. 180
[2] Véase referencia 1 pag. 180

CONTENIDO

RESUMEN

Este manual explica el lenguaje de acceso al Banco de Datos CETENAL (BDC) cuyos programas están escritos en el lenguaje FORTRAN IV.

Una característica de este lenguaje es que dentro de las restricciones que otorga, es ilimitado en su construcción de predicados o preguntas. Se le facilita al usuario su rápido aprendizaje, sin tener conocimientos previos de programación o de computación.

El objetivo del BDC es poder contestar preguntas en forma rápida, tales como: ¿En qué lugares de la República Mexicana se cosechará mejor el maíz?, ¿Qué partes de la República Mexicana están a más de 50 km de las redes de caminos y carreteras existentes?, ¿Cuántas carreteras nuevas se deben diseñar y por dónde van a pasar?, ¿Qué lugares son apropiados para el turismo?, ¿En qué lugares se pueden colocar zonas industriales?, ¿Qué pueblos tienen más de 1000 habitantes y no tienen medios de comunicación?, etc. Contestar éstas y otro tipo de preguntas en forma manual, sería muy la-

borioso e implicaría pérdida de tiempo.[3]

En este manual se describen las funciones que constituyen el lenguaje del usuario; se incluyen ejemplos y sus salidas gráficas.

También en este manual, como otra fase del BDC, se describen: una serie de programas de "aprendizaje", que procesan y "memorizan" las experiencias desarrolladas en la solución de problemas muy variados, como por ejemplo, la búsqueda de zonas urbanas; el BDC indica, como respuesta al problema, en qué lugares existen zonas urbanas al nivel deseado. Sobre este estudio, los programas "inteligentes" encuentran los parámetros y factores comunes que existen en estos lugares en un radio determinado y los "aprenden"; con estas experiencias busca zonas de posible desarrollo urbano.

Entre más zonas urbanas analice, mayor será la correlación de propiedades comunes existentes y más precisa la ubicación de zonas de posible desarrollo urbano.

[3] Véase referencias 5, 10 pags. 180, 181

INTRODUCCION

El Banco de Datos CETENAL (BDC) es una herramienta para manejo masivo de la información (producida por CETENAL y contenida en las diversas cartas), que describe el medio geográfico y socioeconómico.[4]

La información que contiene el BDC es la que un usuario obtendría de las siguientes cartas:

Geológica

Topográfica

Uso del Suelo

Edafológica

Uso Potencial

Se podrían vaciar al BDC otras cartas tales como la Carta Urbana, la de Climas y la Turística, ya que la estructura del mismo acepta gran variedad de información.

El BDC está diseñado para proporcionar varios niveles de consulta. Esto significa que se puede consultar a diferentes tamaños de

[4] Véase referencias 6, 8 pags. 180, 181

área, lo cual activa el acceso a la información.

La información de las cartas se clasifica en tres tipos:

1) Propiedades superficiales.- Comprenden todas las propiedades que están definidas por áreas, por ejemplo: el área que comprende el suelo aluvial, el área urbana, el área de nopalera, etc, están en el BDC como propiedades de este tipo.

2) Propiedades lineales.- Comprenden todas las propiedades que implican uniones entre determinados puntos o comunicación entre ellos, ejemplo: una carretera, vías de energía eléctrica, telegráfica, telefónica, etc.

3) Propiedades puntuales.- Estas propiedades están definidas por su localización, ejemplo: minas, pozos, puntos de verificación, etc.

Las propiedades superficiales están definidas por su clave y por su cantidad, dada como un porcentaje del área del nivel al que pertenecen. Las propiedades lineales y puntuales, están definidas por el número de ellas dentro del nivel al que pertenecen y también por su clave. Hay otras propiedades puntuales, como los servicios de la población y los servicios propuestos para la población en un pueblo, que también están en el BDC.

Como se puede observar, toda la información, ya sea geográfica o socioeconómica, puede ser clasificada dentro de los tres tipos de propiedades anteriores.

El vaciado de la información al BDC se realiza en la Oficina de Bases de Datos del Departamento de Informática de CETENAL con una metodología bien definida.

Las formas de salida son de dos tipos: mensajes escritos por impresora de la computadora, y mensajes por medio de pantallas de CRT, o sea por una terminal remota; esto permitirá tener acceso al BDC, aún estando distante de él. En este manual se describen ambas formas de salida.

Después de definir las funciones del lenguaje de usuario, se describe la forma de combinarlas.

En la parte última del manual se describen los programas de "aprendizaje" del BDC, con una serie de ejemplos para su análisis y conclusiones.

Un aspecto importante del método de "aprendizaje", es el de poder mejorar las experiencias anteriores y agregar nuevos parámetros y factores al BDC.

En los ejemplos de "aprendizaje", también se desarrolló un programa que verifica la confiabilidad del "aprendizaje".

CAPITULO I

DEFINICION DE LOS NIVELES DE CONSULTA

Para establecer los distintos niveles de consulta, se comienza por definir el nivel de regiones. Una región comprende el área dada por un rectángulo de 6° de longitud x 4° de latitud. El territorio nacional está cubierto por 20 regiones aproximadamente. Los siguientes niveles se establecen por subdivisiones sucesivas del rectángulo de menor tamaño que se tenga definido. Así, el nivel de subregiones se forma al subdividir cada región en 12 subregiones que son los elementos de un arreglo matricial de 3 x 4.

Las microregiones constituyen un arreglo matricial de 6 x 4 sobre una subregión. Cada microregión constituye una carta 1:50 000 de las producidas actualmente por CETENAL. La República Mexicana está cubierta por 2336 de estas cartas, aproximadamente. La información vaciada al BDC es extraída actualmente de estas cartas, aunque se vació también una carta urbana escala 1:5 000 y se ha proyectado vaciar las cartas de climas escala 1:500 000. Las microregiones se subdividen en 5 x 5 localidades o grupos y éstos en 4 x 3 sublocalidades o celdas. La figura 1.1 aclara los niveles establecidos y la tabla 1.1 re sume estos niveles y sus características.

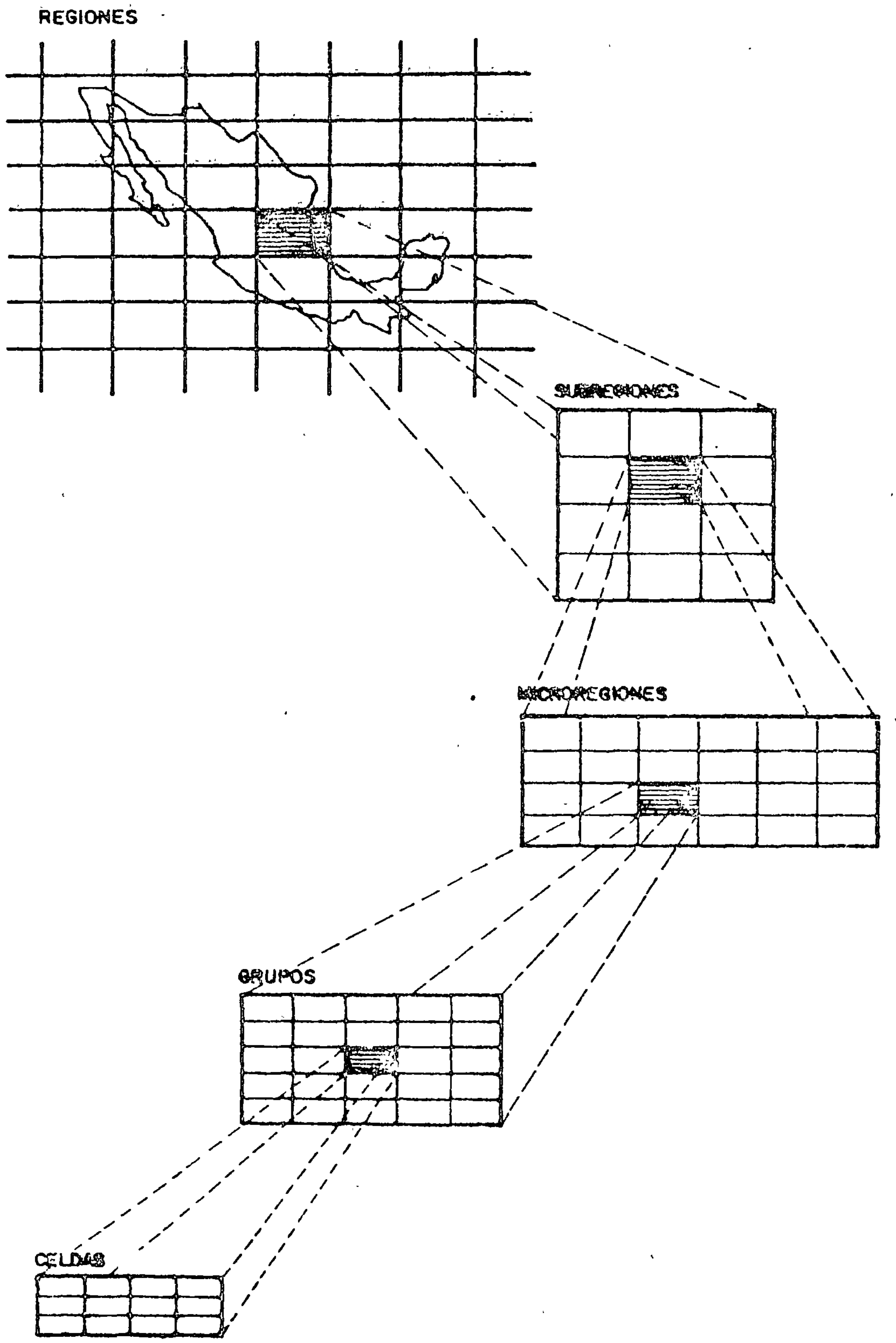


Figura 1.1 NIVELES DE CONSULTA (2a versión)

tabla 1.1 NIVELES DE CONSULTA (2a versión)

NIVEL	TAMAÑO		AREA APROX KM²	CONTIENE	EQUIVALE A
	LONG	LAT			
REGIONES	6°	4°	288 000.0	3 X 4 SUBREGIONES	12 CARTAS 1:250 000
SUBREGIONES	2°	1°	24 000 0	6 X 4 MICROREGIONES	1 CARTA 1:250 000
MICROREGIONES	20'	15'	1000 0	5 X 5 GRUPOS	1 CARTA 1:50 000
GRUPOS O LOCALIDADES	4'	3'	34.7	4 X 3 CELDAS	
CELDAS O SULOCALIDADES	1'	1'	2.9		

Esta estructura de varios niveles agiliza las consultas y además cada subdivisión aumenta la precisión[5].

Una versión más antigua del BDC y que aún se aplica a las zonas de Celaya, Querétaro, Cortázar, Apaseo el Alto y Ojo Caliente, utiliza los niveles dados en la tabla 1.2 e ilustrados en la fig. 1.2[6].

tabla 1.2 NIVELES DE CONSULTA (1a versión)

NIVEL	NOMBRE	AREA APROXIMADA	CORRESPONDE A
4	SUBCUADROS	6.25 km²	
3	CUADROS	25 km²	4 SUBCUADROS
2	ZONA	1000 km²	48 CUADROS O, 192 SUBCUADROS
1	REGIONES	72 000 km²	72 ZONAS O,3456 CUADROS O, 13824 SUBCUADROS
0	REPUBLICA MEXICANA	2 500 000 km²	32 REGIONES O,2336 ZONAS O, 112128 CUADROS O, 448512 SUBCUADROS

5 Véase referencia 7
6 Véase referencias 9, 11

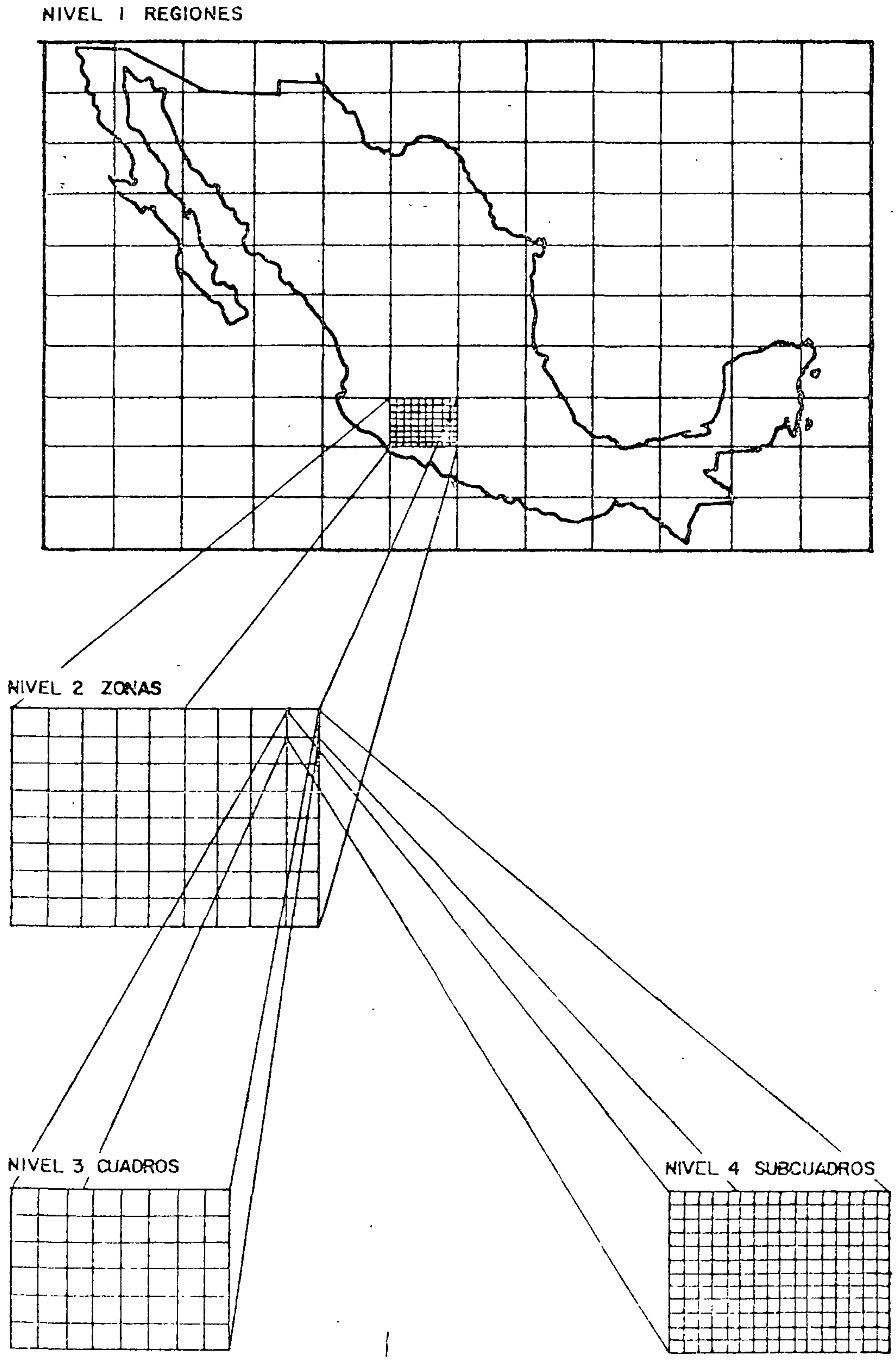


figura I.2 NIVELES DE CONSULTA (1ª versión)

Dentro del BDC a una carta se le llamó zona. Una zona tiene una área aproximada de 1 000 km^2; un nivel superior a éste, en área, es lo que en el BDC se llamó región y es el área comprendida por 72 zonas en un arreglo matricial de 9 x 8 zonas; esta división también aparece en las cartas de avance de CETENAL; cada región tiene una área aproximada de 72 000 km^2 y corresponde a 2° de latitud y a 3° de longitud. La República Mexicana queda contenida en un arreglo matricial de 12 x 10 regiones. Una zona se dividió en 48 cuadros, y un cuadro tiene una área aproximada de 25 km^2; cada cuadro se dividió en 4 subcuadros y cada subcuadro tiene una área aproximada de 6.25 km^2.

CAPITULO II

FORMAS DE SALIDA DE LOS RESULTADOS

Para explicar la forma de salida de los resultados por impresora, se hace referencia a resultados obtenidos con la primera versión del BDC que utiliza los niveles de zonas, cuadros y subcuadros. El significado de estas explicaciones puede hacerse extensivo a las impresiones de resultados obtenidos con la segunda versión del BDC aplicable al Estado de Aguascalientes, y que utiliza los niveles de microregión, localidad y sublocalidad. La única diferencia está en que los distintos arreglos matriciales de cuadros tienen distintas dimensiones.

Como respuesta a la búsqueda a un nivel determinado, la máquina produce en la impresora arreglos matriciales de cuadros de dos tipos: unos de cuadros claros y oscuros, y otros de diferentes tonos de gris (16 en total). Con el primer tipo de arreglos se puede distinguir en qué cuadros las preguntas son verdaderas y en qué cuadros son falsas[5].

[5] Véase figuras 2.1, 2.2, 2.3 pags. 18, 20

Los cuadros oscuros indican en ese lugar la pregunta se hizo verdadera, y los cuadros claros indican que en ese lugar la pregunta se hizo falsa.

Por ejemplo, con un predicado llamado "FRIJOL" se pregunta al BDC en qué zonas hay FRIJOL. Para esto se define la siguiente instrucción:

CALL, BUSCA (FRIJOL, 2, 79)

Y como respuesta se obtiene el resultado que muestra la figura 2.1 en la cual se ve un arreglo de matriz de 2 x 2 zonas. El área de una zona es igual al área de una carta CETENAL escala 1:50 000; las zonas corresponden a la región # 79.

Las zonas oscuras marcan los lugares en donde hay FRIJOL; en estos casos el predicado FRIJOL se cumple. Las zonas claras indican que en esas zonas no hubo FRIJOL, por lo tanto, el predicado buscado no se cumplió en esos casos; o lo que es lo mismo, que el conjunto de propiedades definido por el predicado FRIJOL no se encuentra en lugares donde aparecen zonas claras, (después se verá cómo definir al predicado FRIJOL). Resumiendo: zona oscura significa que sí hay lo que se busca y zona clara significa que no hay lo buscado.

Al buscar el predicado FRIJOL a nivel 3 (cuadros), se obtienen los resultados que muestra la figura 2.2. Esto permite tener una información más precisa que la del nivel 2, ya que dentro de este nivel, que en este caso es la carta de Celaya, se ve con más exactitud cuáles son los cuadros de esta zona que cumplen con el predicado FRIJOL; así se puede seguir haciendo a un nivel más pequeño, como se muestra en la figura 2.3, que corresponde al nivel 4 (subcuadros).

En síntesis, se resume que el predicado FRIJOL se hace verdadero en la zona 1, correspondiente a Celaya, de la región 79. A nivel de cuadro la función FRIJOL se hace verdadera en los cuadros: 5, 6, 13, 14, 17, 21, 22, 25, 26, 27, 28, 29, 30, 33, 36, 37, 40, 41, 42 y 45; esto permite tener más información que en el nivel anterior. Si se desea ampliar esta información se busca a nivel de subcuadro y se ve en qué subcuadro de cada cuadro se hace verdadero el predicado FRIJOL. Los subcuadros que cumplen la función FRIJOL son: los subcuadros **c** y **d** del cuadro 5, los subcuadros **b** , **c** y **d** del cuadro 6, el subcuadro **a** del cuadro 13, los subcuadros **b** y **d** del cuadro 14, los subcuadros **c** y **d** del cuadro 17, los subcuadros **c** y **d** del cuadro 21, los subcuadros **a** , **b** y **c** del cuadro 22, los subcuadros **a** , **b** y **c** del cuadro 25, los subcuadros **a** , **c** y **d** del cuadro 26, los subcuadros **c** y **d** del cuadro 27, los subcuadros **c** y **d** del cuadro 28, los subcuadros **a** , **b** , **c** y **d** del cuadro 29, el subcuadro **a** del cuadro 30, los subcuadros **a** y **c** del cuadro 33, los subcuadros **a** y **b** del cuadro 36, los subcuadros **a** , **c** y **d** del cuadro 37, los subcuadros **a** y **b** del cuadro 40, los subcuadros a y **c** del cuadro 41, el subcuadro **c** del cuadro 42 y el subcuadro **a** del cuadro 45. La tabla 2.1 explica detalladamente la nomenclatura usada[6].

Es preciso hacer notar la importancia que tienen los niveles de consulta en la estructura del Banco de Datos CETENAL.

Esto es, en el ejemplo de la función FRIJOL se buscó a nivel 4, qué subcuadros cumplieron con esta función.

Si no existieran los diferentes niveles de consulta, se tendría que buscar la función FRIJOL en aproximadamente 448 512 subcuadros, lo cual no sería eficiente ni práctico, aún desde el punto de vista del procesamiento por computadora. En cambio, usando los niveles de consulta, se pueden detectar rápidamente los subcuadros que cumplen la función FRIJOL en la forma siguiente:

1. Se busca a nivel 1 regiones[7] de toda la República Mexicana y los resultados se presentarán en un arreglo de matriz de 12 x 10; supóngase que solamente en la región 79 se cumplió la función.
2. Se busca a nivel 2 (zonas en la región 79) y se obtiene la figura 2.1, que muestra la zona 1 oscura; por lo tanto, la función

[6] Véase tabla 2.1 pag. 24

[7] Este nivel actualmente no se ha establecido debido a que el BDC todavía no cuenta con más de 72 cartas vaciadas para formar una región.

es verdadera en ese lugar.

3. Se busca en la zona 1 a nivel de subcuadro y se obtiene la figura 2.3.

Como se puede observar, con los pasos anteriores se obtuvieron en forma rápida los subcuadros que poseen la función FRIJOL, ya que cada nivel está constituido por su descriptor,[8] y no es necesario buscar en lugares donde no existe FRIJOL.

La figura 2.4 muestra una impresión de resultados para el nivel de sublocalidades de una microregión de Aguascalientes.

El segundo tipo de arreglos, formados por cuadros de diferentes tonos de gris, se describe en el capítulo VI.

[8] Descriptor es el conjunto de información que existe en el BDC para cada nivel, véase referencia 1.

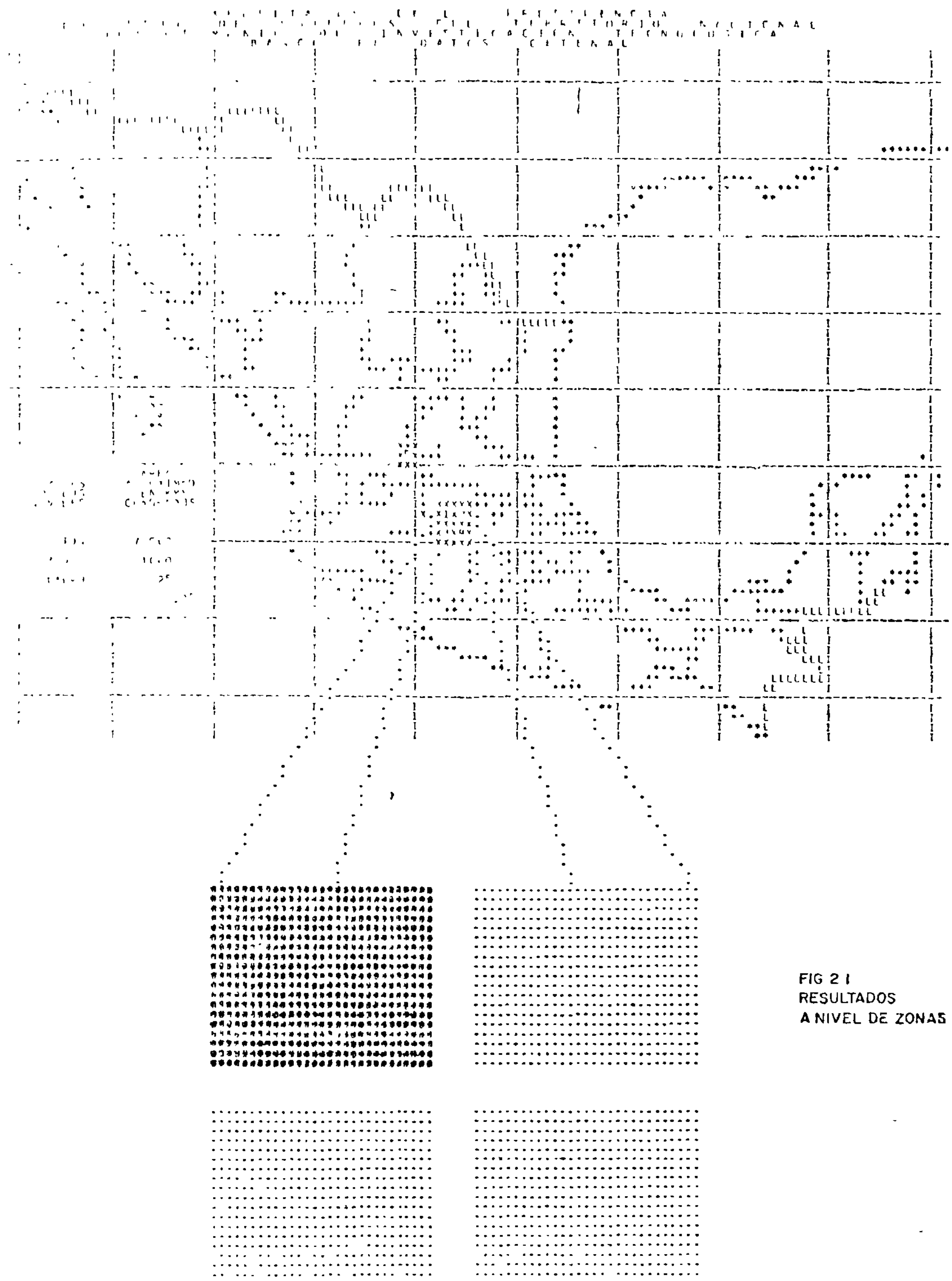


FIG 2 I
RESULTADOS
A NIVEL DE ZONAS

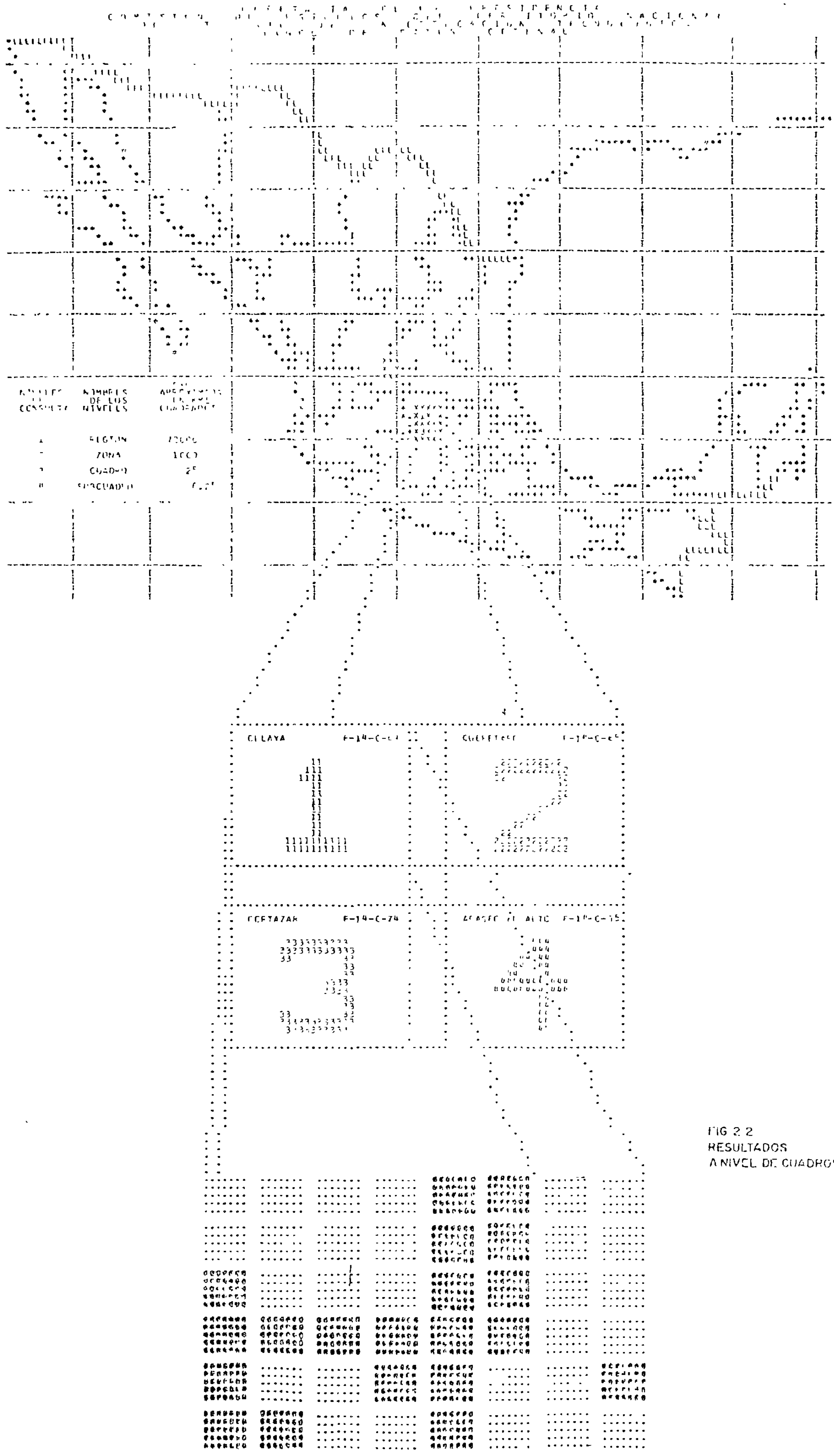


FIG 2.2
RESULTADOS
A NIVEL DE CUADROS

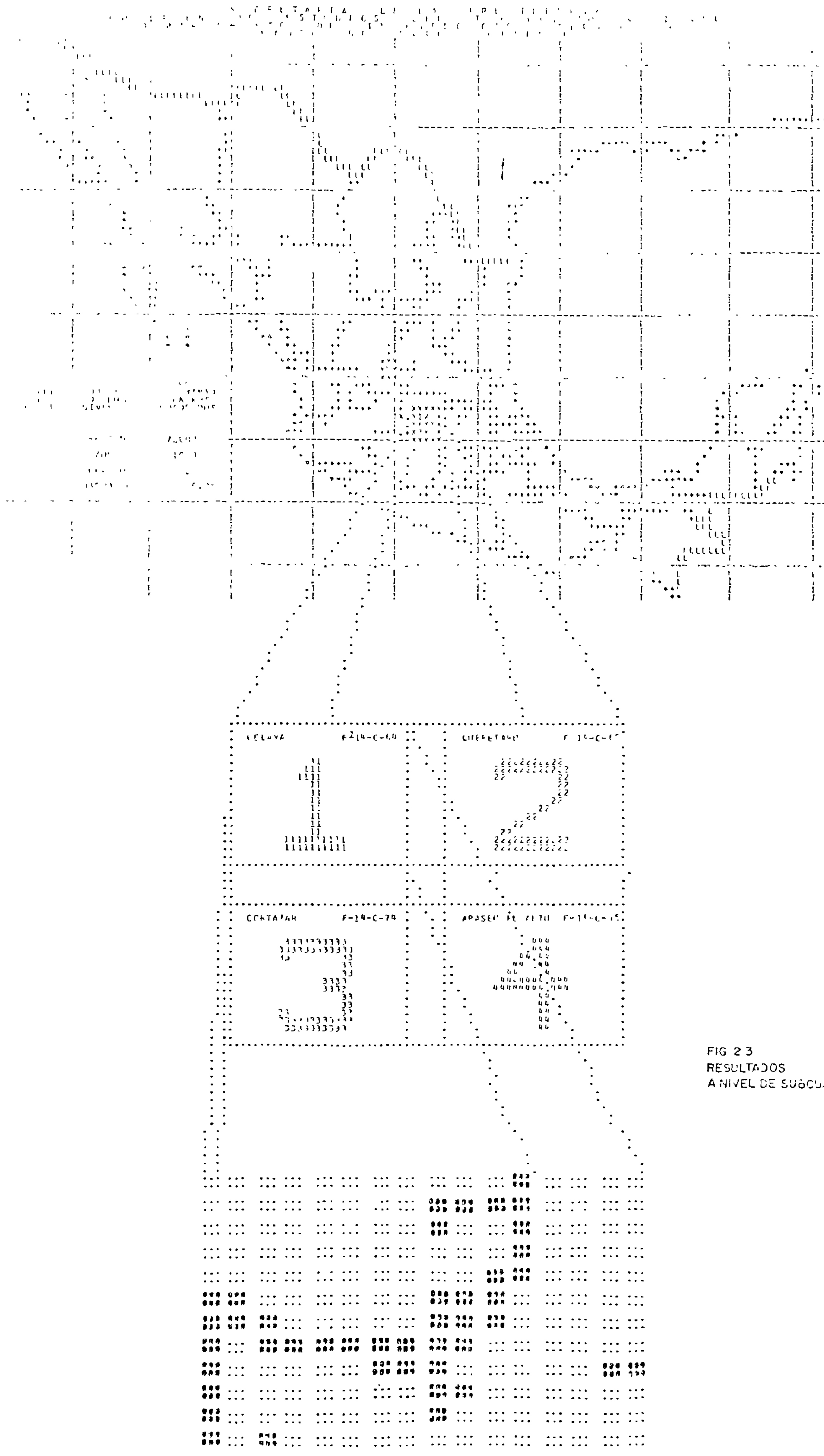


FIG 2 3
RESULTADOS
A NIVEL DE SUBCUADROS

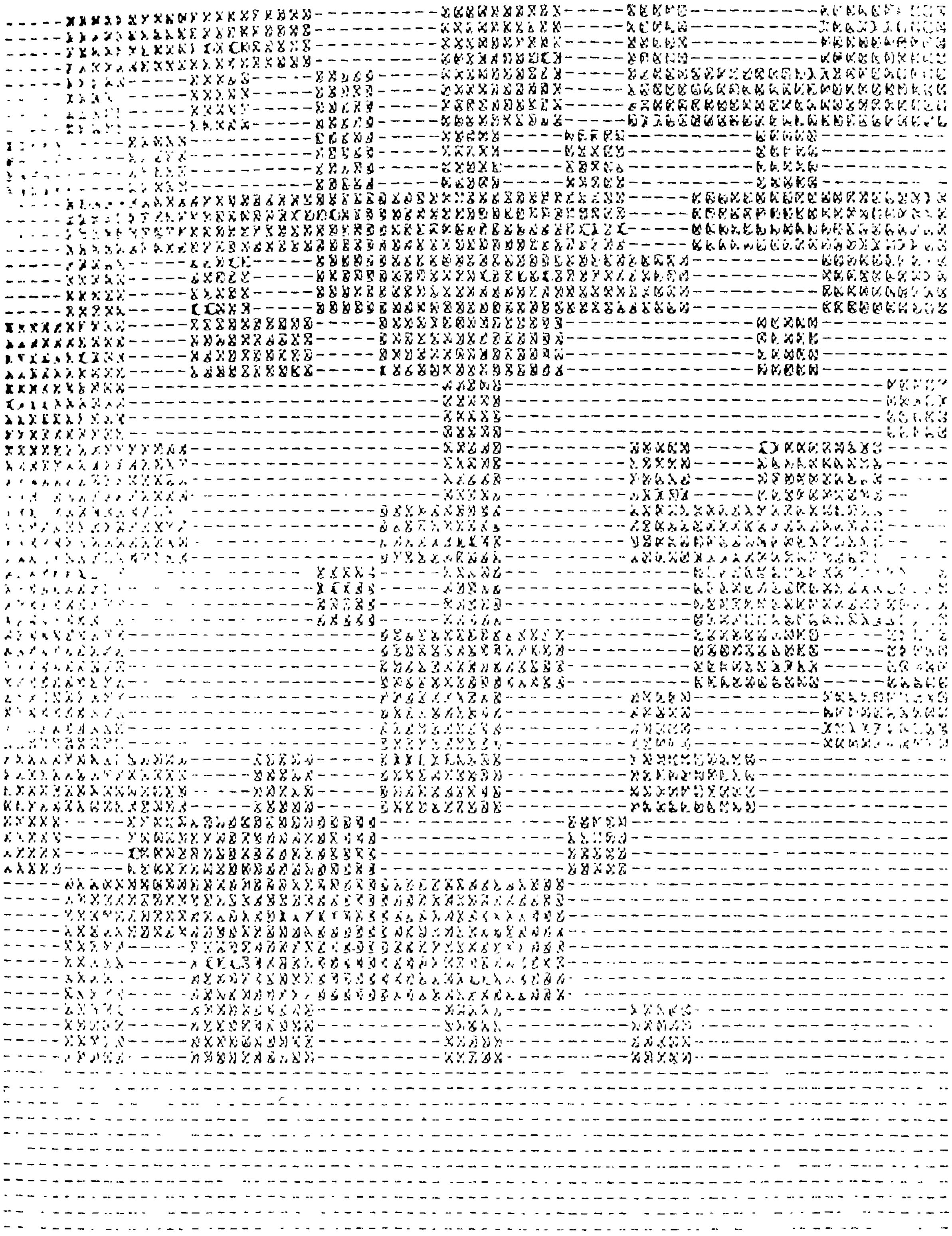


FIG. 24 RESULTADOS A NIVEL DE SUBLOCALIDAD

TABLA 2.1

NOMENCLATURA DE ZONAS, CUADROS Y SUBCUADROS

ZONAS

1	2	1. Celaya	2. Querétaro
3	4	3. Cortázar	4. Apaseo el Alto

CUADROS

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48

SUBCUADROS

ab	ab	ab	ab	ab	ab	ab	ab
cd	cd	cd	cd	cd	cd	cd	cd
ab	ab	ab	ab	ab	ab	ab	ab
cd	cd	cd	cd	cd	cd	cd	cd
ab	ab	ab	ab	ab	ab	ab	ab
cd	cd	cd	cd	cd	cd	cd	cd
ab	ab	ab	ab	ab	ab	ab	ab
cd	cd	cd	cd	cd	cd	cd	cd
ab	ab	ab	ab	ab	ab	ab	ab
cd	cd	cd	cd	cd	cd	cd	cd
ab	ab	ab	ab	ab	ab	ab	ab
cd	cd	cd	cd	cd	cd	cd	cd

En las siguientes secciones se describe el modo de consulta al BDC por medio de una terminal con pantalla de CRT.

Se cuenta con las siguientes funciones adaptadas a teleproceso.

PRO (CCCC, (MN, MY, IG), N)

PROP (CCCC, EN, N, M)

UNADE (CCCC, CCCC, (MN, MY, IG), N)

HAYVIA (CCCC, CCCC)

SERPRO (CCCC, N)

PUEBLO (CCCC)

SERPOB (N, (Y, O), S)

HABIT (N, N)

donde:

CCCC indica la clave utilizada según la tabla apropiada.

(MN, MY, IG) significa alguna de las funciones relacionales.

MN=MENORQ

MY=MAYORQ

IG =IGUALQ

Se utiliza sólo una de ellas.

N es un número que varía de 0 a 99999, según la pregunta.

EN=ENTRE

(Y, O) significa alguno de los operadores lógicos.

Y AND

O-OR

Se utiliza sólo uno de ellos.

S Indica las subdivisiones de los servicios de la población y será un número que incluya a los dígitos 1, 2, 3, 4, 5, 6, 7 sin repetir ninguno según la equivalencia:

a = 1

b = 2

c = 3

d = 4

e = 5

f = 6

g = 7

Creación de una pregunta con base en las funciones anteriores.

- Cada función se construye con base en las especificaciones anteriores, sin incluir espacios intermedios.

PRO (6001, MY, 0) correcto

PRO (6001, MY, 0) incorrecto.

- La posición de cada función no tiene restricción.

- Los espacios en blanco entre funciones no tienen ningún efecto.

- Se utilizan los operadores lógicos Y, O, NO que son equivalentes a:

Y = AND

O = OR

NO = NOT

- Pueden usarse libremente los paréntesis.

- El número máximo de funciones es de 60, inclusive.

- El tamaño máximo de la pregunta es de 10 renglones en la pantalla.

Por medio de las teclas adecuadas se transmite la instrucción al nivel deseado, y se obtiene como resultado una matriz de caracteres, donde * = cuadro oscuro y - = cuadro claro con los significados explicados anteriormente.

CAPITULO III

EJEMPLOS INTRODUCTORIOS

En el resto de este manual se usa la 1a. versión del BDC para describir y ejemplificar el lenguaje de usuario, y los programas de aprendizaje. En general, la 2a. versión funciona de igual manera, salvo pequeñas diferencias que se señalan cuando es necesario. Se utilizaron los cinco tipos de cartas (Edafológica, Topográfica, Geoló gica, Uso del Suelo y Uso Potencial), de cada una de las siguientes zonas; Celaya, Querétaro, Cortázar y Apaseo el Alto. La información que de estas cartas fue vaciada al BDC está descrita en las tablas del apéndice,[9] y corresponde aproximadamente a la información que un usuario extraería guiándose por las leyendas y explicaciones que aparecen en los márgenes de ellas.

Las funciones que dan acceso al BDC permiten describir los sitios (zonas, cuadros o subcuadros), que poseen determinada propiedad o cumplen con cierto predicado.

[9] Véase apéndice pag. 144

Se procede a dar algunos ejemplos sencillos, con el fin de mostrar la simplicidad de su uso. Cualquiera de las dudas que surjan en estos ejemplos, serán aclaradas en secciones posteriores. Por ejemplo, si se desea encontrar, a nivel 3, los cuadros en Celaya que poseen una superficie cultivada mayor del 40%, se define el predicado CULTI[10] mediante el postulado que define a CULTI como todo aquel lugar donde la propiedad 6104 sea mayor de 40, o sea, mayor del 40%. La tabla 1 del apéndice indica que la propiedad 6104 equivale a cultivos.

Una vez hecha esta definición, se procede a buscar CULTI a nivel de cuadro en Celaya, como sigue:

```
CULTI = PRO (6104, MAYORQ, 40)
CALL BUSCA (CULTI, 3, 1)
```

En la segunda versión, la segunda instrucción es como sigue:

```
CALL BUSCAS (CULTI, 1, MICROREG, 2, 3)
```

Las dos instrucciones anteriores, necesitan de ciertas declaraciones e instrucciones adicionales. Los programas completos son: programa que define el predicado CULTI.

```
LOGICAL FUNCTION CULTI (N)
LOGICAL PRO, MAYORQ
```

[10] El nombre dado no tiene importancia pero no debe pasar de 6 letras. Es conveniente usar nombres mnemónicos.

```
EXTERNAL MAYORQ
CULTI = PRO (6104, MAYORQ, 40)
RETURN
END
```

y programa principal que busca CULTI a nivel 3 en la zona 1, esto es, cuadros en Celaya, como se muestra en el listado 3.1.

```
LOGICAL CULTI
EXTERNAL CULTI
CALL BUSCA (CULTI, 3, 1)
STOP
END
```

El resultado se muestra en la figura 3.1 en la cual, se observan cuadros oscuros que son los que tienen más del 40% de su superficie cubierta por cultivos.

Observaciones.- Las declaraciones LOGICAL en los programas definen a CULTI, MAYORQ y PRO, como funciones que adquieren valores lógicos que pueden ser verdaderos o falsos. La (N) de CULTI (N) es una variable muda; las declaraciones EXTERNAL definen a CULTI y MAYORQ como nombres de funciones; las RETURN y END en el predicado, y STOP y END en el programa principal, terminan correctamente los programas. El usuario no necesita entender todo esto; sin embargo, es conveniente apuntarlo.

```
LOGICAL FUNCTION CULTI?NE
LOGICAL PRC,MAYCRQ
EXTERNAL MAYCRQ
C SE DESEA SABER QUE CUADROS DE LA ZONA DE CELAYA TIENEN MAS DEL
* 40 POR CIENTO DE SU SUPERFICIE CULTIVADA
CULTI=PRC?6104,MAYORQ,40E
RETURN
END
```

```
LOGICAL CULTI
EXTERNAL CULTI
CALL BUSCA?CULTI,3,1E
STOP
END
```

LISTADO 3.1 ¿QUE CUADROS DE LA ZONA DE CELAYA TIENEN MAS DEL 40% DE LA SUPERFICIE CULTIVADA?

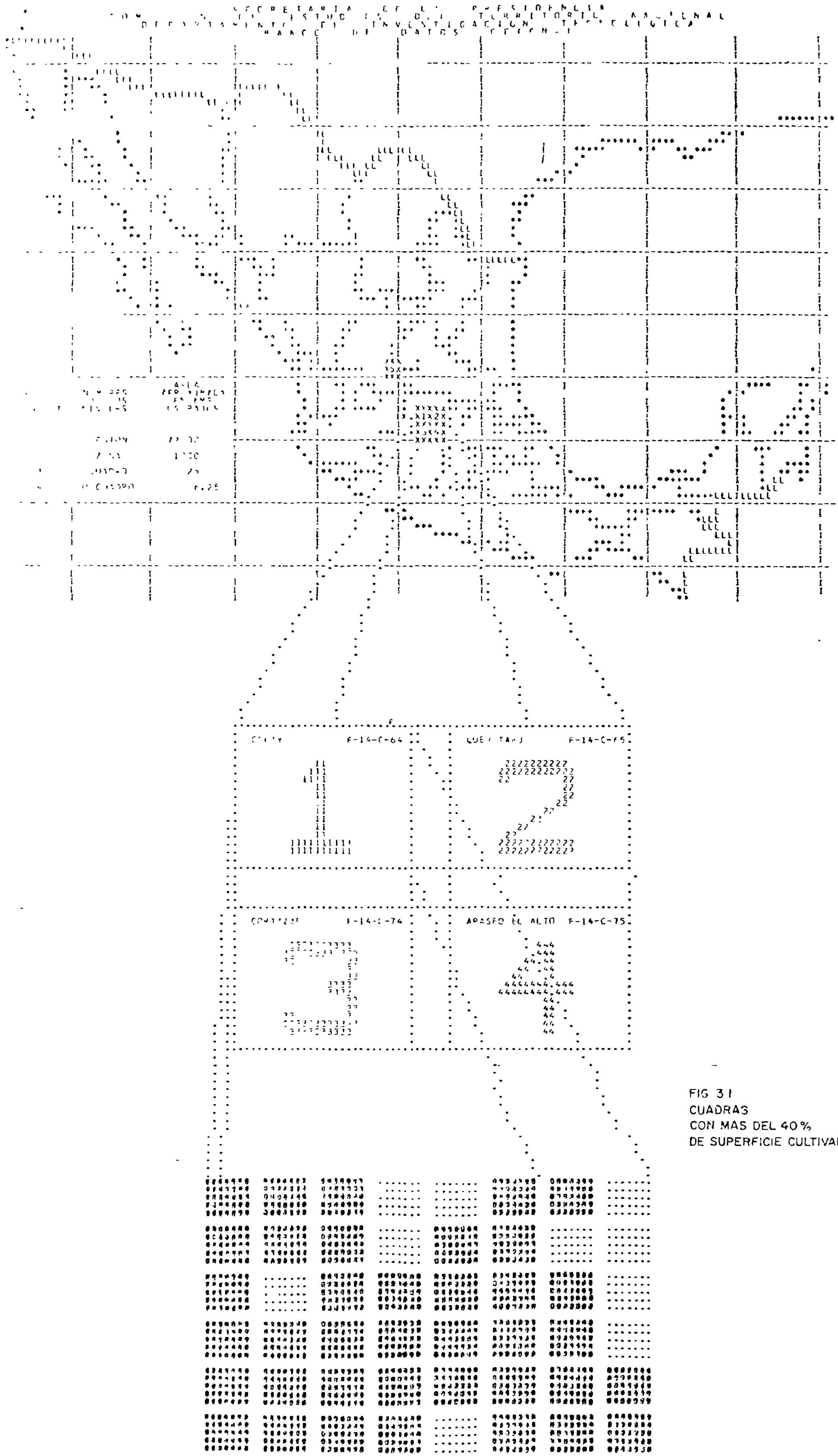


FIG. 3.1
CUADRAS
CON MAS DEL 40%
DE SUPERFICIE CULTIVADA

Para buscar el mismo predicado CULTI a nivel de subcuadro, se dice la siguiente instrucción en el programa principal:

CALL BUSCA (CULTI, 4, 1)

Si hay interés por zonas en las que más del 60% de su superficie está cultivada o sean pastizales, propiedades 6104 y 6201, respectivamente, se dice:

ZONA = PRO (6104, MAYORQ, 60) .OR. PRO (6201, MAYORQ, 60)

Si se quiere hallar zonas que no tengan chaparrales, propiedad 6509, se dice:

NOCHAP = .NOT. PRO (6509, MAYORQ, 0)

Es decir, PRO (6509, MAYORQ, 0) define lugares donde sí existen chaparrales, y con un .NOT. anterior a la expresión, se niega esta propiedad.

El BDC, puede contestar también preguntas como las siguientes:

- ¿Qué lugares tienen pocos álamos y muchos pirules ? Si se conviene en que poco signifique menos del 15%, y muchos signifique más del 70%, la función es:

ALAPIR = PRO (6312, MENORQ, 15) .AND. PRO (6319, MAYORQ, 70)

- ¿Qué lugares están comunicados por cualquier tipo de carretera ? Si se usa la tabla 1 del apéndice,[11] se observa que las pro

[11] Véase apéndice pag. 144

ACKNOWLEDGMENTS

Andrew Clement and T. Peucker gave the triangular idea; Renato Barrera contributed to the concept of a hierarchy of triangles and other good advices. Abel Carreño and Angel García Amaro, of CETENAL, gave good photogrammetric advice. T. Radhakrishnan kindly revised the manuscript.

Work herein reported was partially done under the Joint Research Agreement (IX-1976) between CETENAL and UNAM.

I. INTRODUCTION

The digital representation of three-dimensional surfaces plays important roles in photogrammetry and cartography (drainage patterns, contour lines [5], valleys formation, stereoscopy [4], dimensions of the human body); scene analysis (oclusion of bodies, explanation of regions [6] and objects, range finding, shape from shading [7]); computer graphics (hidden lines, shadows, coloring, specular reflexions); image processing; remote sensing (thickness of ice [8], underground geology), and other disciplines.

The surfaces to be considered are of the following types:

$$z = f(x,y) \quad \text{(I)}$$

$$g(x,y,z) = 0 \quad \text{(II)}$$

Surfaces of type (I) give a single z value ("height") at each pair of coordinates x,y; surfaces of type (II) can represent more general surfaces in the space, for instance the skin of a hand (Figure 1 'Types of Surfaces'). The examples of the paper refer only to surfaces of type (I), but the model hereby proposed can be used for both types.

Our model for a surface consists of a minimum complete cover of triangles; that is, a mutually exclusive, collectively exhaustive finite set of triangles such that (1) each point in the surface is represented by exactly one triangle [12] and (2) every triangle represents at least one point of the surface. To this we add the important restriction: (3) the difference between the coordinates x,y,z of a point in the surface and those x',y',z' of its representative (as given or computed from the model) is less than a pre-specified error or tolerance ε. Thus, for two tolerances $\varepsilon_1 > \varepsilon_2$, the same surface will be represented by two models M_1 and M_2, where M_2 is a refinement of M_1.

For surfaces of type (I), we make x'=x, y'=y and the distance between the real point and the model point is just the difference in heights.

I.1 Non uniqueness of the model

Several complete covers for a surface exist; in order to save memory, it is preferable to use a cover with fewer triangles (hence, each triangle covers in the average a larger area), as long as they obey restriction (3) above. The idea is: (4) not to subdivide a triangle unless it fails to represent reality within the ε tolerance.

Even so, a 3-d surface can be represented by several models complying with (4) above. There is no unique model: by changing the position of the original rectangular frame a new representation is obtained. A different starting rectangle in the procedure or a different orientation of the grid will produce a different cover of triangles.

This is not a problem for 3-d surface representation, as used for instance in applications to cartography and computer graphics. For surface comparison it

TABLE OF CONTENTS

DIGITAL MODEL FOR THREE-DIMENSIONAL SURFACE REPRESENTATION

Dora Gómez and Adolfo Guzmán

Computer Science Dept., IIMAS
National University of Mexico

ABSTRACT

A tree of planar or spherical triangles is used to represent a 3-d surface; rather flat regions will be represented by large triangles, while abrupt zones will require further subdivision of the model into smaller triangles. Their vertices are not placed on a regular grid; they are allowed to fall at (or near) places such as ridges and peaks, where the change in slope is significant.

Starting from a collection, not necessarily good or complete, of "significant" points, the model selects five of them to form four triangles. Each triangle either matches the surface within a prespecified error tolerance, or else is further subdivided, by selecting appropriate "significant" points, into four triangular sons, which then receive in turn the same treatment. The tree stops growing when all the surface is represented within the specified tolerance. The model consists of the vertex points arranged into a table suitable for quick retrieval and interpolation.

Thus, the model guides its own construction; its components (points) are taken from the set of "significant" points, not in an arbitrary fashion but only where and when needed. Since the model proposes the approximate location of the next point to be included in it, the set of "significant" points may be small or non-existant.

A constant signal to noise ratio and a representation thrifty in storage are achieved in this manner.

The model is being tested for use in digital representation of terrain elevation. Large savings in memory are expected, when compared to contour lines storage, for instance.

The paper concludes with some comments in favor of the use of this model to describe gray level pictures.

Key words: three-dimensional surfaces; model; representation; description; terrain; digital; picture; gray level; altitude; computers.

BIOGRAPHICAL SKETCH

Adolfo Guzmán teaches Computer Science at the National University of Mexico, and carries out applied research in digital picture processing. He got his B. S. from the Politécnico de México and his Ph. D. from M.I.T. Dora Luz Gómez is a research assistant at the same place. She studied to be an Actuary at the National University. Her interests are in problem solving and picture processing.

Para proponer las minas que se pueden explotar, se necesitan experiencias anteriores. Los programas pueden adquirir estas experiencias de minas ya existentes en cualquiera de las cuatro zonas.

Se construye la función que busca minas, propiedad 2605, de la siguiente forma:

MINA = PRO (2605, MAYORQ, 0)

También se construye la función de "aprendizaje", en la cual incluye una variable de tolerancia; en este caso del 20%. Puede suceder que los programas de "aprendizaje" no encuentren nada debido a que la variable de tolerancia es muy severa, entonces se puede hacer mayor la tolerancia y así encontrar algunos lugares de los que se buscan.

Para buscar minas en CORTAZAR, se da la siguiente instrucción:

CALL BUSCA (MINAS, 3, 3)

A continuación, el aprendizaje de minas se lleva a cabo con la siguiente instrucción:

CALL LEARN (3, 2605)

Con las muestras (de minas) obtenidas en Cortázar se pueden proponer minas en Celaya como se describió anteriormente.

Según lo que se ha visto, también se puede "aprender" de otras zonas, y así tener mayor confiabilidad en los resultados.

También se puede determinar la confiabilidad con que se está tra-

comunicaciones técnicas

Vol. 9

1978 Serie Naranja: Investigaciones No. 167

DIGITAL MODEL FOR THREE-DIMENSIONAL SURFACE REPRESENTATION. +

Dora Gómez *

Adolfo Guzmán **

+ Technical Report PR-78-19 Laboratorio PR. Submitted to the Journal of Geoprocessing Elsevier Publishing Co.

* Ayudante de Investigador del IIMAS

** Investigador del IIMAS

Recibida: 25/IV/78

INSTITUTO DE INVESTIGACIONES
EN MATEMATICAS APLICADAS
Y EN SISTEMAS
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

APARTADO POSTAL 20-726
MEXICO 20, D. F.
548 54-55

piedades de la 801 a la 810 se refieren a diferentes tipos de carreteras. Con una de ellas es suficiente para comunicar el lugar por donde pasan. La función es:

COMUN = UNADE (801, 810, MAYORQ, 0)

Esta función se interpreta de la siguiente forma: al menos una de las propiedades 801 a la 810 debe ser mayor que cero, es decir, debe existir. Si se quieren conocer todos los lugares que están comunicados y que tienen agua almacenada, en presa o bordo, se dice:

COMAGU = UNADE (801, 810, MAYORQ, 0) .AND.
UNADE (401, 408, MAYORQ, 0)

Otras de las preguntas son:

- ¿Qué zonas se encuentran entre 1600 y 1800 metros sobre el nivel del mar?

RESUL = PROP (101, ENTRE, 1600, 1800)

- ¿En qué lugares hay palmares que ocupan el 25% de la superficie, y están a más de mil metros sobre el nivel del mar?

PALMAR = PRO(6501, MAYORQ, 25) .AND.PRO (106, MAYORQ, 1000)

- ¿Crecerán los nopales sobre suelos de tipo Gléyico?

Se define en que: "crecen los nopales" significa que más del 30% del suelo está cubierto de nopalera, propiedad 6507, y que un suelo es de tipo "Gléyico", cuando la superficie ocupada por la propiedad 3319, suelo Gléyico, es mayor de 70%. Entonces:

NOPG = PRO (6507, MAYORQ, 30) .AND.PRO (3319, MAYORQ, 70)

Se pueden hacer preguntas arbitrariamente largas. En el ejemplo que está a continuación, se buscan lugares de atractivo turístico; éstos pueden ser lugares donde haya cataratas, propiedad 601, o bosques, propiedades 6301 y 6302, o ríos, propiedad 503 y que estén en lugares no muy altos, esto es, a menos de 2 500 m snm; o bien aquellos lugares altos y con volcanes propiedad 2553 o lugares en donde haya más de 4 manantiales termales propiedad 2604, que estén bien comunicados por algún tipo de camino o carretera, propiedades 801 a 810 y que se hallen en terrenos no montañosos, propiedades 5802 y 5803.

Se puede proceder por partes y definir:

UNO = PRO (601, MAYORQ, 0) .OR. UNADE (6301, 6302, MAYORQ, 0)

.OR.PRO (503, MAYORQ, 0) .AND.PRO (105, MENORQ, 2500)

DOS = PRO (106, MAYORQ, 2500) .AND.PRO (2553, MAYORQ, 0)

TRES = PRO (2604, MAYORQ, 4) .AND.UNADE (801, 810, MAYORQ, 0)

.AND. .NOT.UNADE (5802, 5803, MAYORQ, 0)

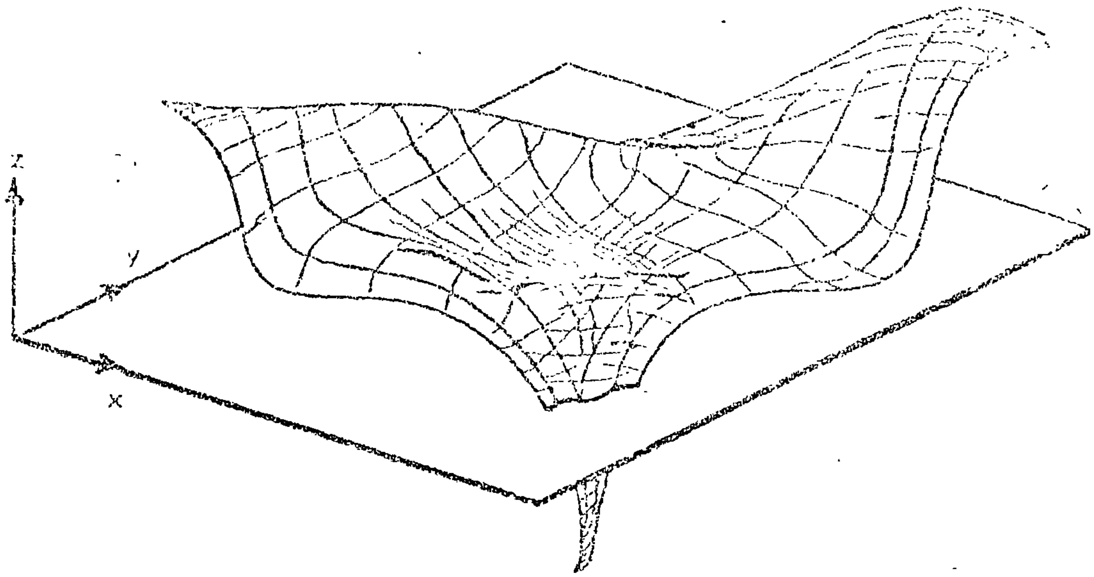
Finalmente se dice:

TURIS = UNO. OR. DOS. OR. TRES

Otro ejemplo es el siguiente:

- ¿En qué lugares de Celaya se pueden explotar minas ?

En este ejemplo, se usan las funciones y rutinas de "aprendizaje".



I. z= f(x,y)

II. g(x,y,z)= 0

FIGURE I 'TYPES OF SURFACES'

The model described in this paper is able to represent either (I) single height surfaces, or (II) more general surfaces.

is much better to have a unique (canonical) model, perhaps through a normalization procedure.

The set of significant points. Using a method external to the model, for instance stereoscopy [4], gradient extraction [10], river following, or others, an initial set of "significant" points is chosen on the 3-d surface that we want to represent. A point is called "significant" if in its neighborhood the change in slope is large.

The model begins by using some of these points; if it later finds necessary to grow, it indicates the approximate place (x,y coordinates) where a new "significant" point should be added to the model.

The model thus consists of a subset of "significant" points, defining a triangular irregular mesh; if the original set of "significant" points is too small, the model will suggest where to add one; if too many, most of them will be ignored (not included in the model); if the procedure that implements "significance" is noisy or unreliable, the model still guarantees the ε tolerance, but storage economy suffers.

Therefore, in a computer implementation, it is not necessary to obtain first the set of significant points and then to pick the model from them; instead, the model can begin to grow as soon as five or six are found, and the procedure that extracts significant points is called by the model as it deems necessary.

1.2 Obtention of the three-dimensional surface

It is assumed that the surface to be modelled already was obtained and exists available in some suitable representation, v.gr., a 2-d matrix containing height values. This data could have been obtained by stereocorrelation /4/ of a pair of pictures, by interpolation of digitized contour lines /1/ or by other means.

II. CONSTRUCTION OF THE MODEL

In order to describe the model, it is necessary to explain

(1) its constituent parts. In this case, they are vertices ("significant" points from the 3-d surface to be represented) that form planar, but tilted, triangles.

(2) how the model is stored; the data structure used to keep the model in memory (primary or secondary storage). A tree of triangles, each with none or four sons, is used.

(3) the use of the model: the procedure to follow for reconstruction of the

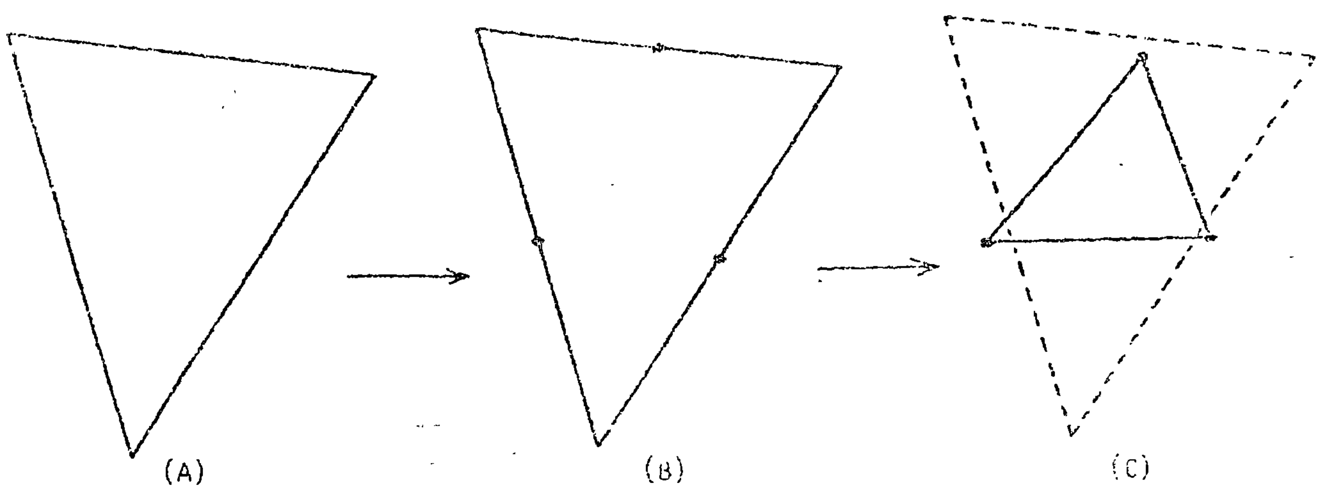


FIGURE 2 'TRIANGLE REFINEMENT'

If it is necessary to refine triangle (A), three new vertices are proposed at the mid-points (B) of the sides; "significant" points are located near those mid-points; once they are found (C), four new triangles stand instead of the original (A).

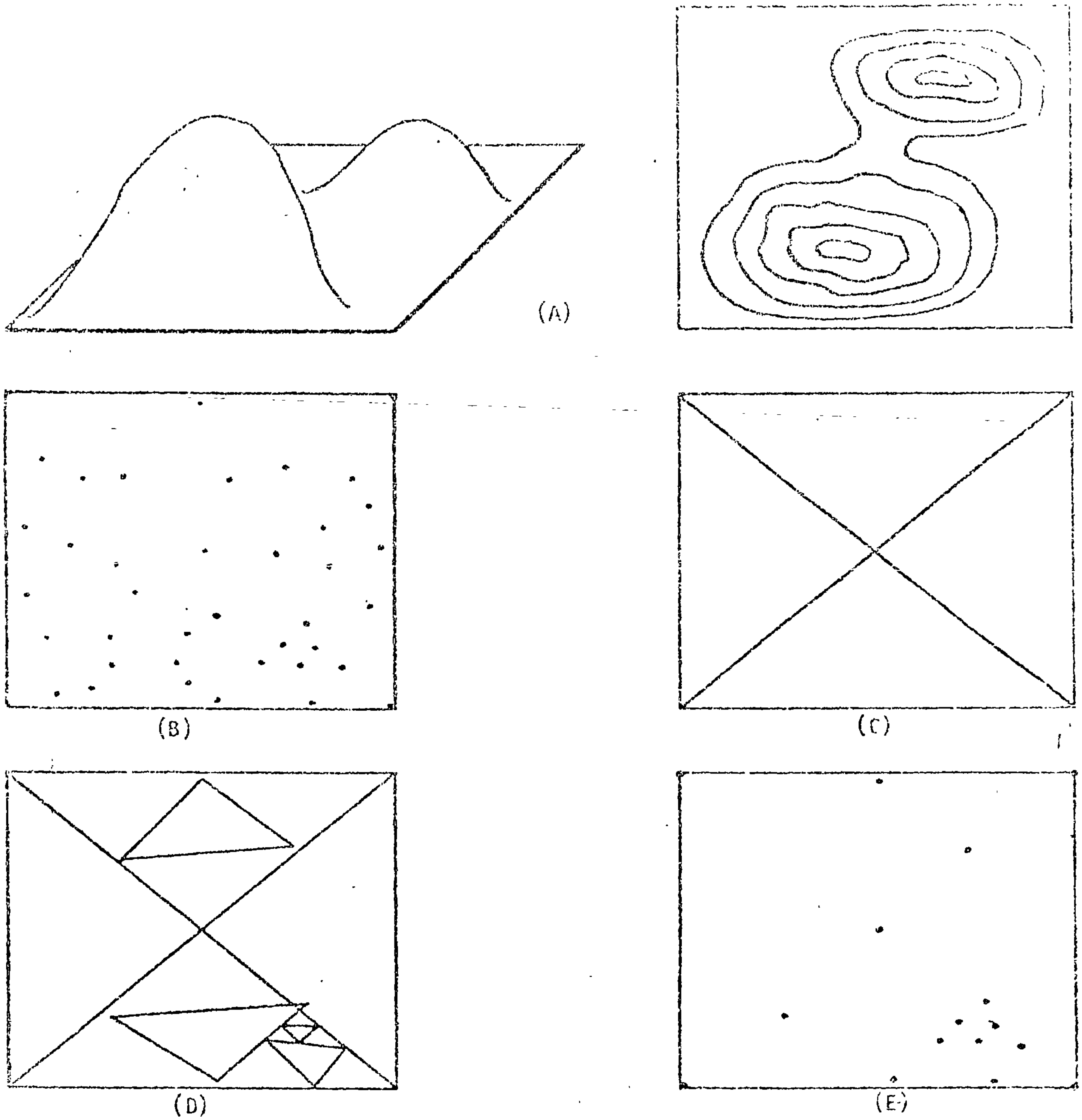


FIGURE 3 'MODEL BUILDING'

(a) the surface. (b) the "significant" points.
(c) the four initial triangles.
(d) the final triangles.
(e) the points of the model, usually a subset of (b).

Triangles (d) are in the space (they project out of the paper); similarly, points (e) have three coordinates.

3-d surface from the model; the way to obtain the coordinates of a point on the surface from the cover of triangles. Here a directed access is used to the correct triangle starting from the top of the tree of (2), and falling down the appropriate chain of triangle sons, using little search and no backtracking.

(4) the construction of the model, i.e., the obtention of its parts from the 3-d surface. A recursive procedure will be presented, where the model guides its own construction, by suggesting places (x,y coordinates) where to incorporate into itself points from the 3-d surface that are also "significant" with respect to changes in slope.

II.1 The parts of the model

To represent a surface $z = f(x,y)$, the model uses a collection of planar tilted triangles; each of them is defined by its three vertices, chosen to lie on the surface $z = f(x,y)$ to be represented.

All the points inside the triangle are interpolated linearly: the surface inside the triangle is considered flat (but not horizontal, in general). Since the real surface $z = f(x,y)$ is not flat, an error is introduced by this assumption. If everywhere in the triangle this error (height difference) does not exceed a tolerance ε, the planar triangle is considered to be a good (and final or "terminal") representative for that region of the surface, and it is included in the model. If the error is larger, the triangle is discarded by dividing it into four smaller triangles, each of which in turn undergoes the same treatment.

Initially the surface is divided into a small set of arbitrarily chosen large triangles; if the surface is bound by a rectangle (as it is frequently the case in maps), four triangles are chosen as shown in part C of Figure 3 'Model Building.'

The final model contains triangles (of different sizes) that represent the surface $z = f(x,y)$ with a tolerance ε. Each of the vertices of these triangles was proposed by the model by dividing a triangle in four through inclusion of new vertices near the middle points of the sides (Figure 2 'Triangle Refinement.').

Once every triangle is refined, the vertices ((E) in fig. 3 'Model Building') are stored in an appropriate way, suitable for quick data retrieval for surface reconstruction.

II.2 When to stop refining

A triangle such as (A) in figure 2 'Triangle Refinement' is refined further, unless

1) the difference between the real height $z = f(\bar{x},\bar{y})$ and the computed height $\bar{z}$ at the center of mass $(\bar{x},\bar{y},\bar{z})$ of the triangle is smaller than

ε, and

2) every point in a grid of points spaced at most K units apart [illegible] (A) is within ε of the real point on the surface $z = f(x,y)$.

Test (1) is a quick test; test (2) is applied only if (1) does not [illegible] a difference exceeding ε.

K, the distance between two points in the grid of (2), is a function of ε, normally, $K = \min(\varepsilon/m, K_o)$, where m is the mean slope of the surface at the triangle (A), and K_o is the diameter of the smallest topographic feature (hill, ravine) that it is necessary to represent in the model. Generally K_o is given by the user of the model: "be sure to check the model every 500 horizontal meters for accuracy"; then $K_o = 500$.

Flow diagram. The procedure for construction of the model could be summarized as:

- Let T be the set of triangles that are candidates to be included in the model. Initialize T with the four triangles of (C), Fig. 3 'Model Building.'

- Mark every triangle of T as "terminal" if it passes tests (1) and (2) of Section II.2 'When to stop refining.' If these tests fail for a triangle, mark it "non-terminal", divide it into four sons (cf. Fig. 2 'Triangle Refinement') and add them to T.

- Exit when all triangles of T (including all the additions to T) are marked (either "terminal" or "non-terminal.") Then T is the model.

BEGIN

T the four initial triangles of fig. 3C;

For every triangle in T

if it passes tests (1) and (2) of Sec. II.2
then mark it 'terminal'
else mark it non terminal and
add its four sons to T;

END.

TABLE 1 'FLOW DIAGRAM FOR MODEL BUILDING'

This simple program constructs surface models such as that shown in fig. 6 'Model Example.'

A non-terminal triangle is not needed in the model, since

(1) its accuracy is worse than ε, and

(2) some of its descendants are *a fortiori* terminal triangles, hence suitable for modelling.

Thus, the model could be just the collection of terminal triangles.

This is advisable when the cover is made of similar triangles (q.v.), where it is easy to pick up the correct triangle for surface reconstruction. If the triangles are not similar, it is preferable to retain the non-terminal triangles into the model. This facilitates the addressing of the correct terminal triangle that gives the height $\bar{z}$ of a point (x,y) (i.e., the point $(x,y,\bar{z})$ that represents the point (x,y,z) of the 3-d surface). More of this in the section IV 'Data Retrieval for Surface Reconstruction.'

II.3 Cover of similar triangles

Two polygons are similar if the corresponding angles are equal, the sides parallel and their length proportional.

If in figure 2 'Triangle Refinement' we stop the refinement at (B), choosing the midpoints as new vertices to include in the model, the final cover of the model is composed of two families of similar triangles, because a line joining the middle points of two sides is parallel to the third side.

A word of caution: the triangles are not similar as they lie in the 3-d space. Their projections on the plane x.y do form a family of similar 2-d triangles (for triangles *a b e* and *c d e* of fig. 4 'Model Visualization' and all their descendants) another family of similar 2-d triangles for triangles *d a e* and *b c e*, and all their descendants.

The advantages of the cover of similar triangles are:

(a) storage of these triangles is easy. [3] sotres a hierarchy of squares.

(b) reconstruction of the surface from the model becomes simplified.

(c) a set of "significant" points (B in fig. 3 'Model Building') is not needed.

The disadvantage comes from (c):

(d) The model might contain more points, since they are not special or significant: they are not the best to choose for interpolation of planes.

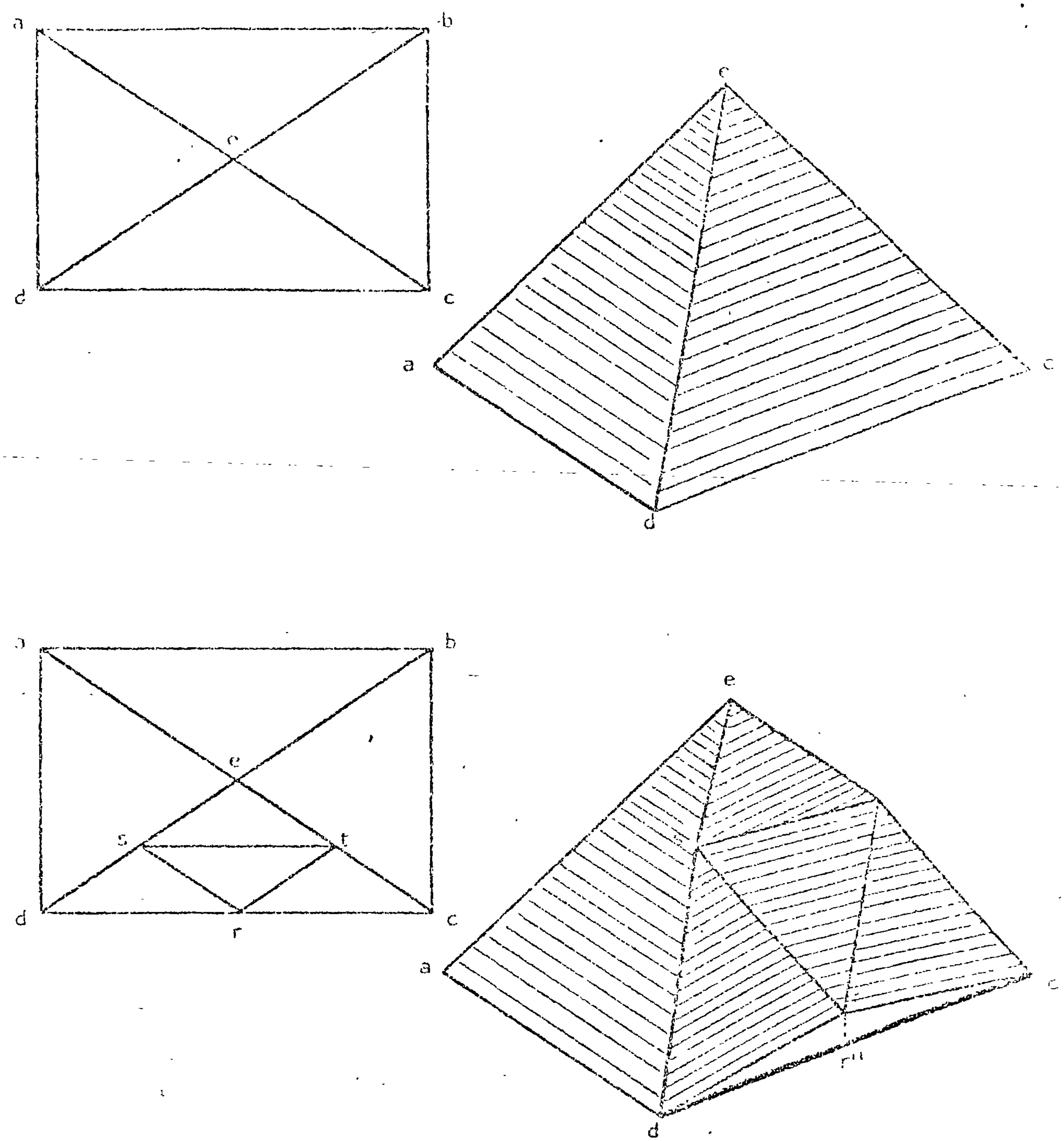


FIGURE 4 'MODEL VISUALIZATION'

We try to give an isometric view of the appearance of the triangular model in 3-d space. Point $S=(s_x,s_y,s_z)$ does not lie on the 3-d line $d \rightarrow e$; but point $(s_x,s_y,0)$ does lie on the line $(d_x,d_y,0) \rightarrow (e_x,e_y,0)$.

Triangles such as d e s or d t e are not part of the model; they represent no part of the real 3-d surface because they lie vertical. They are useless.

CONVENTIONS I. Refer to part (A) of Fig. 5 'Data Structure.'

Vertices of triangles which are sons of the rectangle are named as shown. The correct names for (A) are:

rectangle: a b c d

M = triangle a b e

N = triangle b c e

O = triangle c d e

P = triangle d a e.

CONVENTIONS II. Refer to part (B) of Fig. 5 'Data Structure.'

Vertices of triangles that are sons of triangles are named clockwisely, starting with the vertex that also belongs to the father.

If the triangle to be named is the internal triangle (P), then start with the vertex that falls near the middle point of line 1→2, where 1 is the first of the vertices that belong to the father, and 2 is the second of them.

The correct names for triangles of (B) are:

triangle 1 2 3 (first vertex is 1)

M = triangle 1 4 6

N = triangle 2 5 4

O = triangle 3 6 5

P = triangle 4 5 6

TABLE 2 'NAMING CONVENTIONS'

These conventions are important for correct storage of vertices (such as 4 of the internal triangle P), and its subsequent appropriate retrieval for reconstruction of the 3-d surface. For a use, see definition of procedure 'altitude' in Section IV.

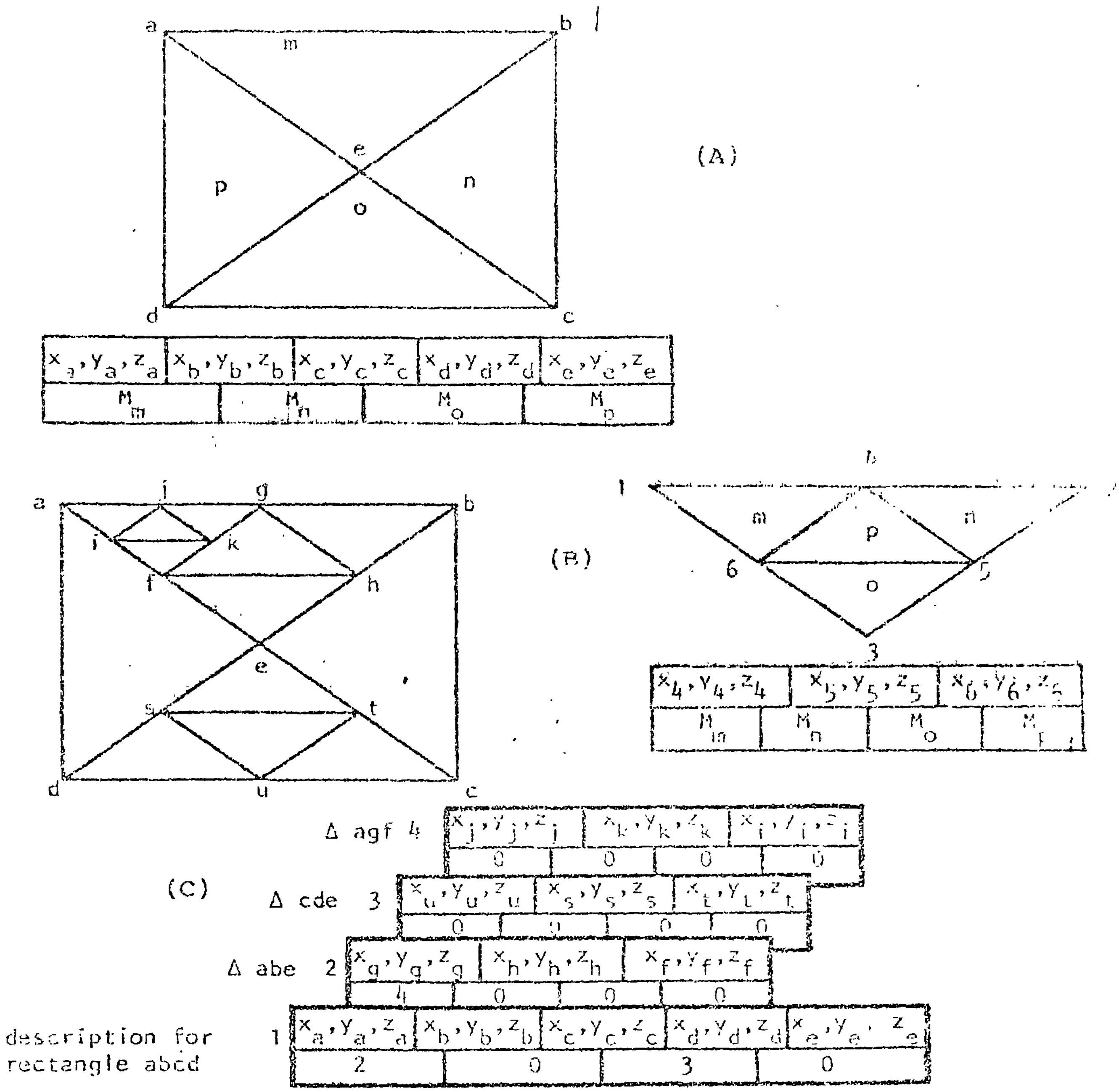


FIGURE 5 'DATA STRUCTURE'

(A) Storage conventions for the initial rectangle.
(B) Storage conventions for a non-terminal triangle *1 2 3*.
Its sons are *M, N, O, P*.
(C) Example of a model and its data structure. Only a non-terminal triangle uses up a frame. The 2 0 3 0 marks of frame 1 mean that son M is non-terminal and it is described in frame 2, son N is terminal, son O is non terminal and it is described in frame 3 and son P is terminal (mark = 0 means terminal). The model is stored in a matrix (C) which is a collection of frames.

III. DATA STRUCTURE FOR MODEL STORAGE

This section describes the way to organize the storage of the model. Essentially, the storage consists of a collection of triangles. Each triangle is stored in a "frame"; each of them contains

- three internal vertices

- a "terminal" or non-terminal mark for each son.

The terminal mark (zero) indicates that a triangle son already fulfills the ε accuracy, hence it (the son) has no sons of its own --need not be further subdivided.-- The non-terminal mark, an integer different from zero, indicates the location (frame) in the model matrix occupied by this triangle son. Thus, when a node is marked as non-terminal, the mark itself also says where (in what frame) that son is stored. See fig. 5 'Data Structure' and Table 2 'Naming Conventions.' Slightly different conventions were used in IIMAS-UNAM /5/.

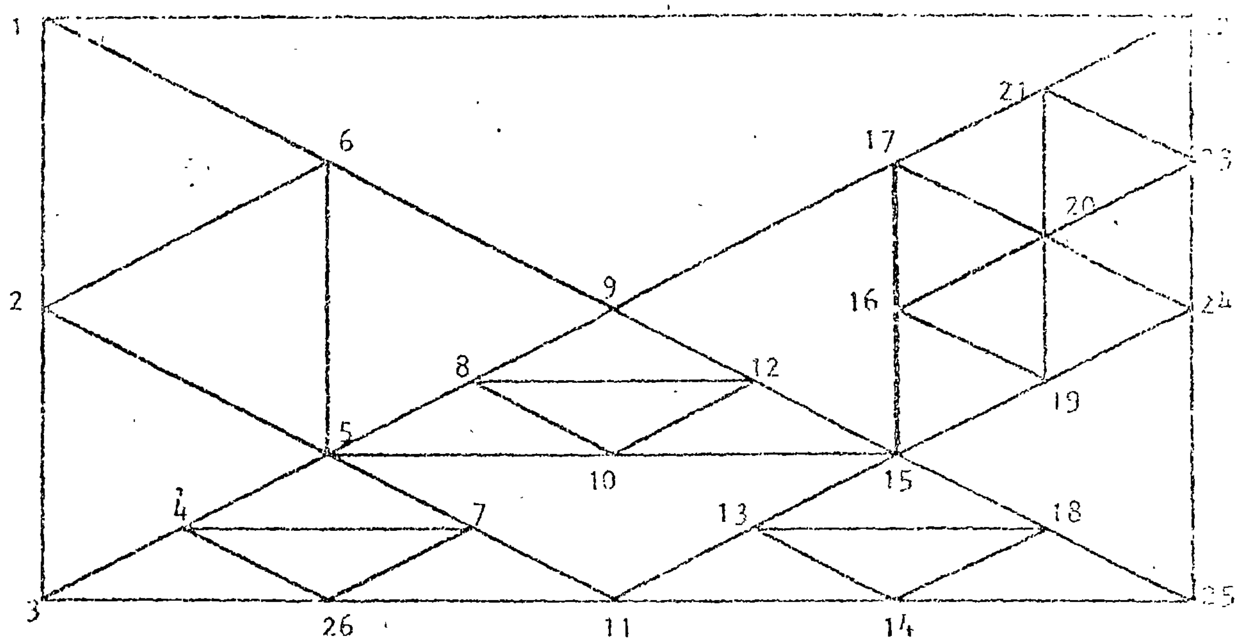
The model is stored in a matrix (C, Fig. 5 'Data Structure') which is a collection of frames. A non-terminal triangle occupies a frame; it stores, clockwisely (B, Fig. 5) its three central vertices and a mark specifying for each son whether it is terminal or not. A terminal triangle does not use a frame, since it has no sons. But a non-terminal triangle could very well have four terminal sons. That is the case of frames 4 to 9 of fig. 6 'Model Example.'

The initial frame, frame 1, is stored in a slightly different manner (part A of fig. 5 'Data Structure') because it describes a rectangle.

A more complicated example is given in fig. 6 'Model Example.'

<u>Storage of vertices</u>. When describing a non-terminal triangle (v.gr., triangle *1 2 3* of B, Fig. 5 'Data Structure'), only vertices 4, 5 and 6 are stored in the frame belonging to that triangle *1 2 3*, since vertices 1, 2 and 3 were undoubtely stored in the ancestors of triangle *1 2 3*. This avoids multiple storage of vertices, and exploits the fact that in order to examine whether a point (x,y) falls inside the 2-d /13/ triangle *1 2 3* or not, we already asked a similar question to the ancestors of *1 2 3*. In this way the coordinates of vertices 1, 2 and 3 are already known when triangle *1 2 3* is accessed (cf. Section IV 'Data Retrieval for Surface Reconstruction').

A vertex is stored by storing its three coordinates x, y, z. Some duplication (not triplication or multiplication) occurs when a vertex such as 5, 20 or 45 in Fig. 6 'Model Example' gets stored by two non-terminal brother triangles. For instance, vertex 5 is stored at frame 4 that describes triangle 3 1 9, and also at frame 3 that describes triangle 25 3 9. The trivial one will be to keep a table of vertices, and to store in the frame pointers to the table. Instead of the three coordinates x,y,z.



	frame	a	b	c	d	e	Δ M	Δ N	Δ O	Δ P
		VERTEX								
(□ 1 22 25 3)	1	1	22	25	3	9	0	2	3	4

		VERTEX 4	VERTEX 5	VERTEX 6				
(Δ 22 25 9)	2	24	15	17	5	0	0	6
(Δ 25 3 9)	3	11	5	15	7	8	9	0
(Δ 3 1 9)	4	2	6	5	0	0	0	0
(Δ 22 24 17)	5	23	20	21	0	0	0	0
(Δ 24 15 17)	6	19	16	20	0	0	0	0
(Δ 25 11 15)	7	14	13	18	0	0	0	0
(Δ 3 5 11)	8	26	4	7	0	0	0	0
(Δ 9 15 5)	9	8	12	10	0	0	0	0

FIGURE 6 'MODEL EXAMPLE'

This example was constructed using the rules (A) and (B) of Fig. 5 'Data Structure' and Table 2 'Naming Conventions.'

Each frame consists of vertices and pointers to other frames. Only non-terminal triangles occupy a frame of the matrix. This matrix <u>is</u> the model.

This table of vertices is not used in our model because it saves little storage:

(1) if both a pointer and a vertex coordinate occupy a word of memory, then to use the table requires 2 pointers + 3 coordinates = 5 words; not to use the table requires 3 coordinates + the same 3 coordinates = 6 words;

(2) if for some reason triangle 25 3 9 selects vertex 5 as the "significant" point near the mid-point of side 3-9 (Refer to Fig 6), but triangle 3 7 9 selects vertex 5' (a different vertex, near vertex 5 but not the same) as the "significant" point near the mid-point of side 9-3, then the table wastes memory.

III.1 Simplified storage for cover of similar triangles

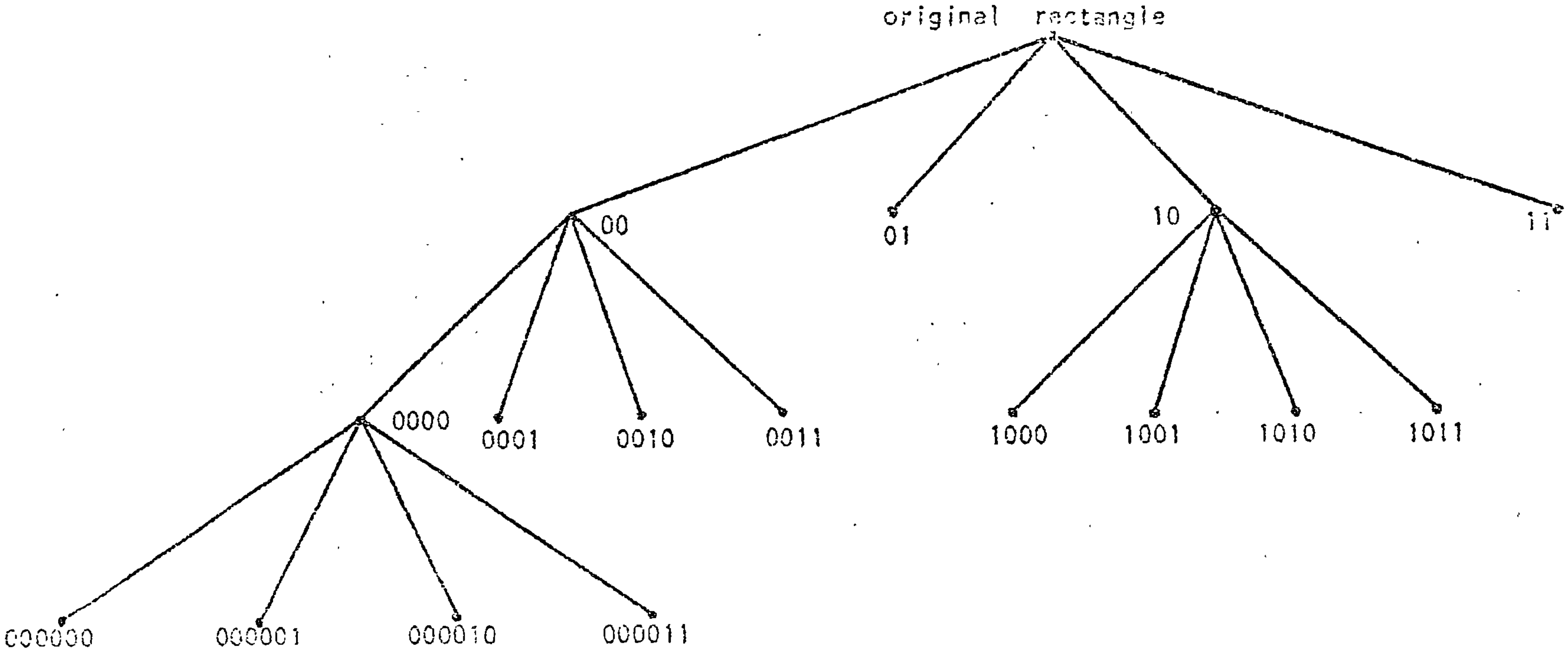
If we assume that the rectangle *a b c d* (Fig. 5 'Data Structure) is a square and that the "significant" points are exactly at the mid-points of the sides of the triangles, instead of near them, then all the two-dimensional triangles [13] are similar (in fact, they are isosceles right angled triangles) and the (x,y) coordinates of any vertex need not be stored, since they are the average of the (x,y) coordinates of the vertices of an appropriate side.

The new representation for square *a b c d* of Fig. 5 is:

	Frame #	vertex a	vertex b	vertex c	vertex d	vertex e	Δ M	Δ N	Δ O	Δ P
(*rectangle a b c d*)	1	Z_a	Z_b	Z_c	Z_d	Z_e	2	0	3	0
			vertex 4	vertex 5	vertex 6					
(*triangle a b e*)	2		Z_g	Z_h	Z_f		4	0	0	0
(*triangle c d e*)	3		Z_u	Z_s	Z_t		0	0	0	0
(*triangle a g f*)	4		Z_j	Z_k	Z_i		0	0	0	0

If the original area is not an square but a rectangle, we will have two families of similar two-dimensional triangles.

If we denote the sons M, N, O and P by 00, 01, 10 and 11, then we could form from figure 5 the following tree:



First generation

Δ a b c = 00

Δ b c e = 01
Δ c d e = 10

Δ d a e = 11

Second generation

Δ a g f = 00 00

Δ b h g = 00 01
Δ e f h = 00 10
Δ g h f = 00 11

Δ c u t = 10 00
Δ d s a = 10 01
Δ e t s = 10 10
Δ u s t = 10 11

Third generation

Δ a j i = 00 00 00
Δ g k j = 00 00 01
Δ f i k = 00 00 10
Δ j k i = 00 00 11

These codes could be combined with the z values to render a compact model. We do not pursue this further. In a similar manner, a tree of squares can be represented [9].

IV. DATA RETRIEVAL FOR SURFACE RECONSTRUCTION

In order to recover the 3-d surface, it is sufficient to ask the model what is the z value for any pair x,y. This is realized by the function ALTITUDE.

```
ALTITUDE (x, y)     % returns the height z of the point (x,y) as obtained from
                    % the model. It is defined as:

a := MODEL[1,1] ;    % first vertex of frame 1. Frame 1 is the rectangle.
b := MODEL[2,1] ;    % MODEL [*,L] is the frame L, a non-terminal triangle.
c := MODEL[3,1] ;    % MODEL [*,*] is the matrix containing the whole model.
d := MODEL[4,1] ;
e := MODEL[5,1] ;
m := MODEL[6,1 ];    n := MODEL [7,1] ;     o := MODEL [8,1] ;
p := MODEL [9,1 ];   %retrieving the pointers to the sons.
error := -1;

ALTITUDE := if  inside (a,b,e,x,y)
            then        if m = 0  then  height (a,b,e,x,y)
                        else ZETA (a,b,e,x,y,m)
            else if inside (b,c,e,x,y)
            then        if n = 0  then  height (b,c,e,x,y)
                        else ZETA (b,c,e,x,y,m)
            else if inside (c,d,e,x,y)
            then        if o = 0  then  height (c,d,e,x,y)
                        else ZETA (c,d,e,x,y,m)
            else if inside (d,a,e,x,y)
            then        if p = 0  then  height (d,a,e,x,y)
                        else ZETA (d,a,e,x,y,m)
            else error;
END ALTITUDE.
```

Function INSIDE (a,b,c,x,y) is true if the point (x,y,0) is inside the triangle $(a_x, a_y, 0)$, $(b_x, b_y, 0)$, $(c_x, c_y, 0)$ with sidewalks. Ref. Fig. 7 'Coplanar Sidewalks.'

A point p is inside triangle a b c if p and c fall on the same side of a b and p and b lie on the same side of a c, and p and a rest on the same side of b c. A thesis [4] contains listings and results.

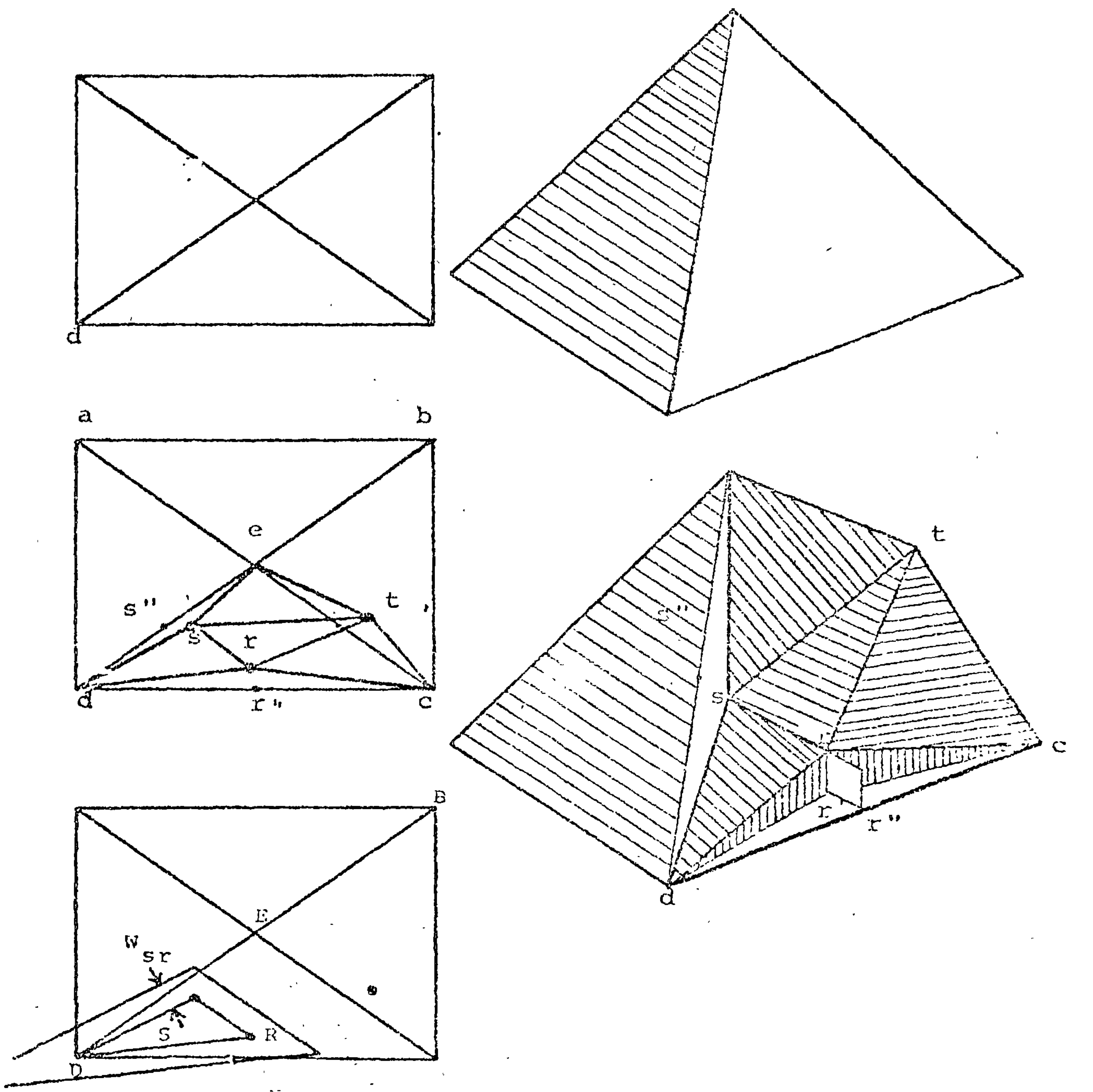


FIGURE 7 'COPLANAR SIDEWALKS'

Compare with Figure 4 'Model Visualization.' If point $r = (r_x, r_y, 0)$ does not fall on line $(c_x, c_y, 0) \rightarrow (d_x, d_y, 0)$, an horizontal area $c\ d\ r'$ will be without coverage by the triangles; a corresponding part of the 3-d surface will fail to be represented. The cure for this is to give "Flaps" to the triangles, so that triangle $r\ d\ s$ (and its other three brothers) are enlarged by a coplanar sidewalk that covers up to r [12].

```
Procedure ZETA (v1, v2, v3, x, y, m) is defined as
v1 := model [1,m] ;
  v5 := model [2,m ];
    v6 := model [3,m] ;
      mm := model [4,m] ;
        n := model [5,m] ;
          o := model [6,m] ;
            p := model [7,m] ;

ZETA := if inside (v1,v4,v6,x,y)
        then      if mm = 0 then height (v1,v4,v6,x,y)
                  else ZETA (v1,v4,v6,x,y,mm)   %see Table 2 'Naming Convention'
        else if inside (v2,v5,v4,x,y)
        then      if n = 0  then height (v2,v5,v4,x,y)
                  else ZETA (v2,v5,v4,x,y,n)
        else if inside (v3,v6,v5,x,y)
        then      if o = 0  then height (v3,v6,v5,x,y)
                  else ZETA (v3,v6,v5,x,y,o)
        else if p = 0
        then      height (v1, v4, v6,x,y)
        else      zeta (v1,v4,v6,x,y,p);
END ZETA.
```

The search for the correct triangle that represents a point generates no backtracking. At each level of the tree of triangles, we simply go down to the next level through the appropriate son (that son containing the point), until we hit a terminal triangle, where we compute the height by a planar interpolation.

V. CONCLUDING REMARKS

Since a gray level picture can be seen as a surface in three dimensions, z being the gray level value, it is in principle possible to use the models described here to represent them. This could have use for shape comparison of these surfaces, but the authors have not experimented with this. The idea, anyway, is to use models with large ε (large error tolerance, coarse representation) to compare two surfaces; if the models are equal (in some appropriate sense, for instance, the quantized z values agree) then we could afford comparison with a smaller ε (more accurate representation). In this way the shape similarity between any two 3-dimensional surfaces (or any two gray level pictures) can be ascertained. A related paper [2] develops this idea fully for two-dimensional flat regions (binary pictures). The largest problem with this approach is to find a normalization procedure (the basic rectangle of [3]) that will produce a unique model for the 3-d case: it is easier to compare canonical models.

The method described in this paper is currently being implemented and tested for representation of topographic surfaces formerly described by their contour lines.

Merging of models into a larger model. If four adjacent surfaces a, b, c, d are represented by models a, b, c, d, the model of the joint surface (a,b,c,d) is formed by creating a new frame I (cf. Fig. 5 'Data Structure') which has as terminal pointers M_a, M_b, M_c, and M_d, pointers to the frames I of a, b, c and d.

Significant points vs. correlation points. The significant points (also called surface-specific points [11]) are those points of the terrain where slope changes in an important way. The points that a correlation routine finds in an easy way, found for instance in the two pictures of a stereo pair, are called "correlation points:" they are points that are easy to correlate in the pictures, because the gray levels in their neighborhood are quite different from others, hence they can be identified rapidly and unmistakably. But they will not necessarily fall on top of "significant" points.

The components of the model. The model so far described and its construction can be seen as formed by:

- a tessellation of polygons [5] (triangles in this case);

- an accuracy criteria, which tells whether a polygon of the model needs further refinement (in our case, comparison of modelled vs. real heights, cf. Section II § 2 'When to stop refining');

- a procedure to refine the model (in our model, select a significant point near the middle point of a side);

- a manner to store the model (as exemplified in Fig. 6 'Model Example');

- a way to access the model (as seen in Section IV 'Data Retrieval for Surface Reconstruction');

- a method to reconstruct the surface from the model (this is given by the procedure *height* (a,b,c,x,y) evaluated at the appropriate triangle $a\ b\ c$ which contains the point (x,y,0); the appropriate definition of containment is embodied in procedure *inside* (a,b,c,x,y), which takes into account, for instance, the "flaps" of Fig. 7 'Coplanar Sidewalks.').

V.3 Suggestions for further work

1. Refer to Fig. 7 'Coplanar Sidewalks.' Do not use k_1 = 10% for the width of the sidewalks. Compute instead the maximum distance that $(r_x, r_y, 0)$ can be from r'' for the enlarged triangle $t\ d\ s$ to meet still the error tolerance. This has to do with average slopes of the triangles.

2. Refer to § III.1 'Simplified storage for cover of similar triangles.' Fully develop the model that uses the representation of each triangle as a string of pairs of binary digits, v.gr., triangle $g\ h\ j$ = 00 00 01 (the son H of the son M of the son M of the rectangle).

3. Do not retrieve the triangles from the root of the tree (cf. Section IV 'Data retrieval for surface reconstruction') but store them so as to access them by a double binary search on the coordinates of the vertices [4].

4. Consider the methods of this paper and of [1,7] as similar procedures that address data representation at arbitrary accuracy levels, and use them for shape comparison.

References

1. Bribiesca, E., and Avilés, R. Codificación en cadenas y técnicas de reducción de información para mapas y dibujos lineales. Informe CCAL-74-7, IBM Latin American Scientific Center (Mexico City), Nov. 1974.

2. Bribiesca, E., and Guzmán, A. Shape description and shape similarity measurement for two-dimensional regions. Submitted to Fourth International Joint Conference on Pattern Recognition, Kyoto, Japan, November 1978.

3. Bribiesca, E., and Guzmán, A. Shape numbers: a notation to describe pure form and to measure resemblance and difference in shape. Report PR-78-15, Computer Science Dept., IIMAS, National University of Mexico. 1978.

4. Gómez, D. Modelos digitales del terreno de precisión variable. B. S. Thesis, Facultad de Ciencias, Universidad Nacional de México (in preparation).

5. Gómez, D. Tessellation of triangles of variable precision as an economical representation for DTM's. Proceedings of the Digital Terrain Models Symposium St Louis, Mo. 1978. Available from American Society of Photogrammetry.

6. Guzmán, A. Analysis of curved line drawings using context and global information. In Machine Intelligence VI, (D. Michie and B. Meltzer, eds) University of Edinburgh Press. 1971. Chapter 20.

7. Horn, B. K. P. Shape from shading: a method for obtaining the shape of a smooth opaque object from one view. Ph. D. Thesis, E. E. Dept., M.I.T. 1970. Project MAC Technical Report MAC-TR-79.

8. Jensen, H. H. Collaboration in P hysics within the Nordic countries. Europhysics news 7, 5, May 1976, pp 1-4.

9. Klinger, A., and Nikitas, A. Picture decomposition, tree data structures, and identifying directional symmetries as node combinations. Computer Graphics and Image Processing (to appear).

10. Signor, G., and Nadler, M. Une application de la corrélation numérique d'images: la stéréophotogrammetrie automatique. Congres AFCET/IRIA. Reconnaissance des Formes et Traitement des Images. Paris 1976.

11. Peucker, T., Fowler, R. J., Little, J.J., and Mark, D. M. Triangulated irregular networks for representing three-dimensional surfaces. Simon Fraser University, Burnaby, Canada. Technical Report # 10. April 1976.

12. The width of the sidewalk is a constant fraction k_1 (say, 10%) of the corresponding median: $w_{sr} = k_1 \cdot \overline{s \to r}$. To incorporate these sidewalks to the model, it was only necessary to modify the definition of the function *inside*, so that a point is "inside" triangle α *d* s if it falls inside it or

at the sidewalk (Refer to fig. 7).

If no significant point r is found near enough r'' so as to meet the 1% requirement, the model orders to find (to fabricate) a new significant point with coordinates x,y closer to r_x'', r_y'': the model only accepts significant points close enough to r'' (r'' is the midpoint of $(c_x,c_y,0) \rightarrow (d_x,d_y,0)$) to ensure that the resultant flaps will not be wider than k_1. An optimal way (suggestion 1) is needed to compute k_1.

The sidewalks slightly contradict the assumption (1), uniqueness of representation for a point, of the Introduction. This is of no importance.

13. Refer to (B) of Fig. 5 'Data Structure.' Triangle 1 2 3 is a triangle in space, a 3-dimensional triangle, since each vertex has three coordinates. But we could project triangle 1 2 3 into the x-y plane, by setting z=0 at each of its vertices. In this manner we obtain a triangle in the x-y plane, which we call the "two-dimensional" triangle 1 2 3.

comunicaciones técnicas

1978 Vol. 9
Serie Naranja: Investigaciones No. 166

SHAPE DESCRIPTION AND SHAPE SIMILARITY MEASUREMENT FOR TWO-DIMENSIONAL REGIONS. +

Ernesto Bribiesca ++

Adolfo Guzmán *

+ Technical Report PR-78-18, Laboratorio PR
Submitted to 4th. International Joint Conference on Pattern Recognition, Kyoto, Japan, Nov. 1978

++ Investigador DETENAL

* Investigador del Depto. de Computación del IIMAS

Recibida: 13/III/78

INSTITUTO DE INVESTIGACIONES EN MATEMATICAS APLICADAS Y EN SISTEMAS

UNAM

SHAPE DESCRIPTION AND SHAPE SIMILARITY MEASUREMENT FOR TWO-DIMENSIONAL REGIONS

Ernesto Bribiesca
Depto. Investigación
CETENAL (Mexico)

and

Adolfo Guzmán
Computer Science Dept., IIMAS
National University of Mexico

Abstract

We analyze the forms or shapes of flat regions limited by simply connected curves. A procedure is given that deduces from every region a unique number (its shape number) independent of translation, rotation and scaling.

The precision in the representation of the shape of a region by one of its shape numbers is indicated by the order of that shape number; high orders are more accurate for shape description. Informally, the number of ternary digits of a shape number will tell its order. The paper contains tables of all the shape numbers of order k, for several k. Nevertheless, these tables are not necessary for computing the shape number of a region. The shape number of any order can be deduced solely from the region; no shape matching, comparison or table lookup is necessary.

We then introduce the degree of similarity between the shapes of two regions and give an algorithm for computing it from the corresponding shape numbers. Two regions with shapes that look alike will have a high degree of similarity. No string matching or grammatical parsing is necessary to find out how close in shape two regions are. Informally speaking, the degree of similarity between the shapes of two regions is the highest optical resolution (power of the magnifying lens) that still confuses them. We then define the distance between two shapes and find it to be an ultradistance.

In this way, a quantitative study of shape is possible.

Finally, a related Theory "B" of shapes is presented that disregards the excentricity of a region and offers additional advantages for shape comparison.

Key words: chain encoding; shape numbers; figure description; discrete shapes; shape representation; similarity of form; shape comparison; distance between forms.

INTRODUCTION

Scene Analysis seeks to understand a scene, for instance by assigning names to its different parts and components as well as by explaining their relations and structures.

Local and global information [3], that is, shape and context, play an important and mutually supporting role in Scene Analysis. If we look at scenes found in coloring books for children (Fig. 1, 'STREET SCENE'), the explanation (name, purpose, role) of each part is derived both from its shape and from the context, that is to say, from the names of the parts close to it.

The role of shape in Scene Analysis

Take Fig. 1 'STREET SCENE' which lacks color, texture, gray levels, and only has shapes, sizes and structure. One can still make a good "explanation" and understanding of it. Consequently, one of the authors has proposed [3] to represent explicitly these three components, for instance by a graph where the nodes contain shape and size information about each region, and the arcs represent different relations ("above," "between," "surrounded by") among the nodes.

It is therefore necessary to be able to describe the shape of an object (part, region); to compare shapes; to decide how close two given shapes are, or what is their resemblance or dissimilarity. A numerical reliable measure for these concepts will give rise to a quantitative study of shape.

Definitions

Region. A simply connected portion of a plane limited by a curve boundary. That is, no holes, no self-intersecting boundary. Closed boundary. The region is uniquely defined by the curve it has as boundary.

This paper deals with shapes of regions, but the shape numbers used here can also be applied to open curves.

Freeman chain in four directions. For a given region and a given square grid of fixed orientation and size, the Freeman chain in four directions is the curve obtained by walking clockwise on the grid (on the "wires"

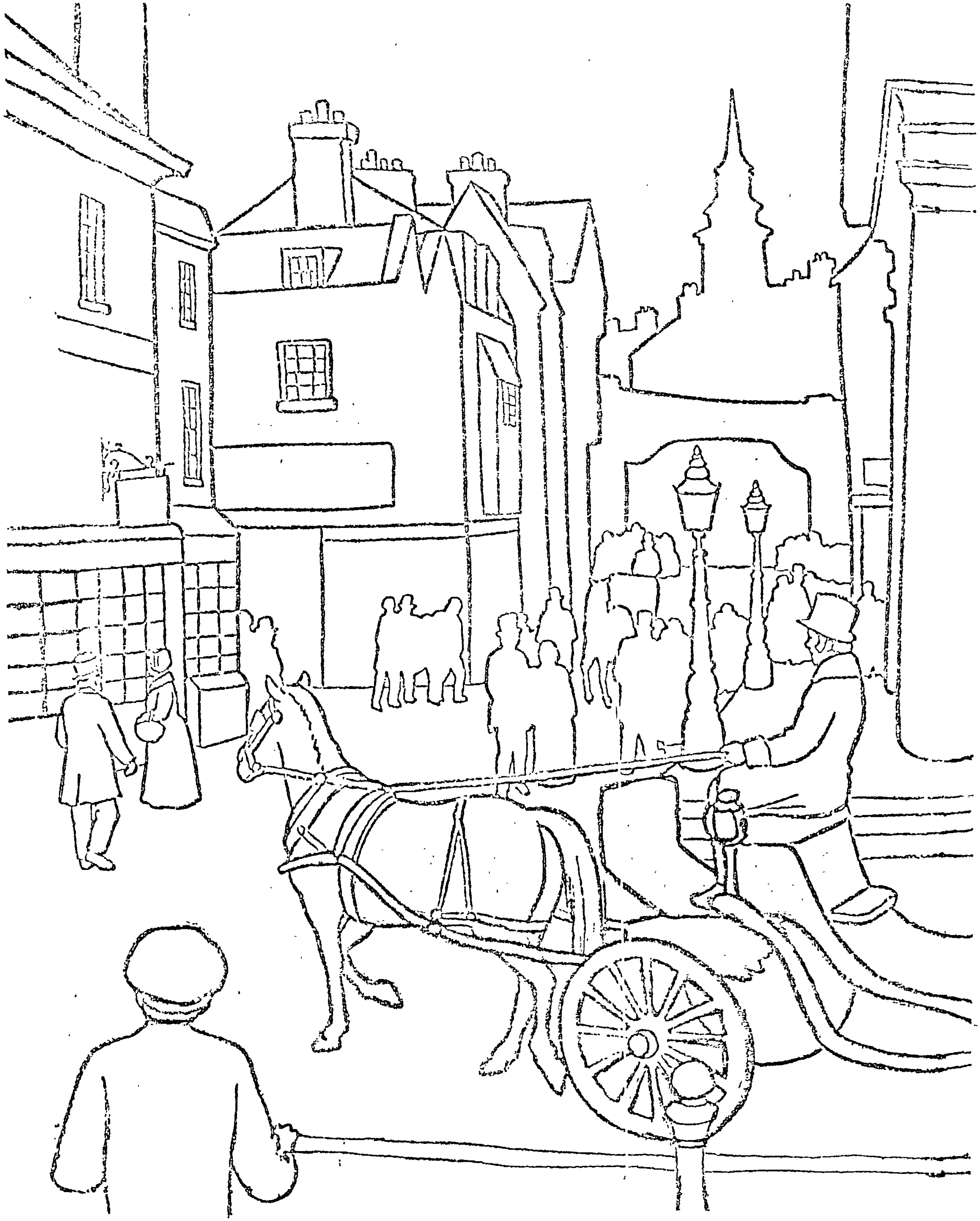


FIGURE 1 'STREET SCENE'

Each closed part (region) of this scene could of the grid) around and outside the squares that contain more than 50 % of the region (Fig. 2).

The chain number (Fig. 2d) is obtained by clockwisely replacing each step along the curve by the number 1, 2, 3 or 4, according to Fig. 2e. See suggestions 1 and 2 at the end of the paper.

Sometimes this procedure will break thin portions of regions and one will end up with two non-connected chains. These are degenerate regions for that grid, which have no shape numbers (q. v.) (Fig. 8.II).

Derivative of Freeman chains. It is the chain number (Fig. 2f) obtained by clockwise replacing each salient (convex) corner of the Freeman chain (Fig. 2c) by a 1, each straight corner by a 2, and each concave corner by a 3, as figure 3g suggests.

The number obtained (Fig. 2f) will be different if we change the size or orientation of the grid.

Major axis of a region. The straight line connecting the two perimeter points furthest away from each other (Fig. 3b).

Ocassionally, there will be more than one major axis in a region. In that case, select that which gives the shortest minor axis.

Minor axis of a region. A segment perpendicular to the major axis, and of length and position such that the box formed by these two axis just encloses the region (Fig. 3a).

Other axis for similar purposes are given in [3], pp 338-342, and in [2].

Basic rectangle of a region. It is the rectangle having its sides parallel to and of sizes equal to the major and minor axis, such that it just encloses the region (fig. 3d).

Excentricity of a rectangle. It is the ratio of the long to the short side: $e \geq 1$.

Excentricity of a region. It is the excentricity of its basic rectangle. It is the ratio of its major to minor axis. This definition coincides with that for an ellipse.

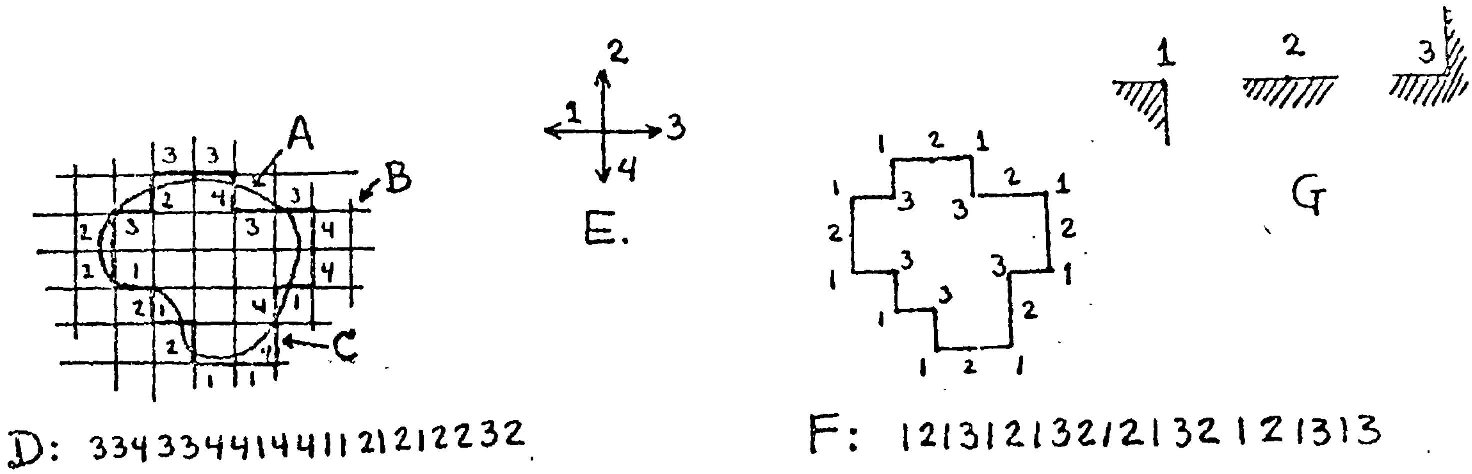


FIGURE 2 ' C H A I N S '

a: the region. b: the grid. c: the Freeman chain in four directions. d: its chain number. e: the four directions of (b) used to code (c) into (d). f: the derivative of (c). g: the three types of corners used to code (c) into (f).

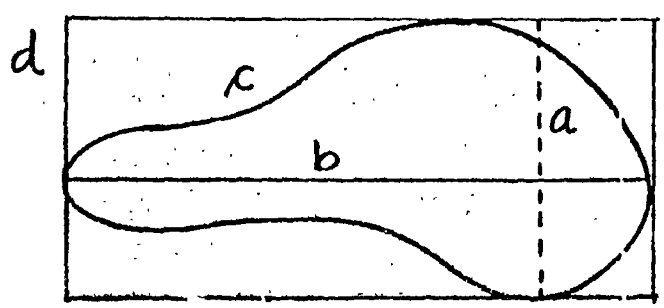


FIGURE 3 ' DEFINITIONS'

a: minor axis of (c). b: major axis of (c). c: region. d: basic rectangle of (c).

THE SHAPE NUMBER OF A REGION

If a notation is going to be used to represent the shape of a region, it has to be independent of the position, orientation and size of such region. It should be reproducible: a region, when translated, magnified and rotated should still give the same description as when untransformed. Two regions with different shapes should produce different descriptions. Finally, the shape number should be unique for a given region; for instance, it should not depend on an arbitrary starting point or a particular coordinate system.

If the notation can be deduced exclusively from the region, without comparison with a table of canonical shapes or shape descriptors, for instance, then we can expect savings in memory and computer time for the procedure that finds out the shape description.

In this section, we first produce finite families of shape descriptors (every member of a family has the same <u>order</u>); we then exhibit a way to find out, for an arbitrary region, its shape descriptor of any order. This descriptor indeed qualifies as a notation to represent shape.

In the next section we will see that this descriptor also permits to measure the similarity or analogy between the shapes of two regions.

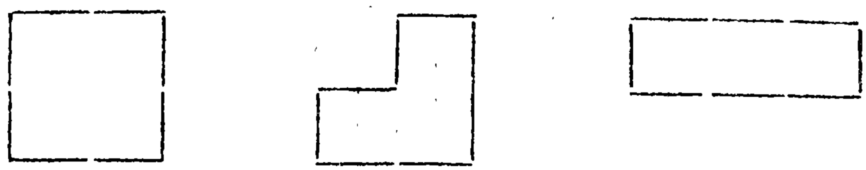
Discrete shapes

Regions of special interest are crated when it is required to form a closed curve using o sticks of the same length, but joining them end to end either colinearly or forming 90 ° corners. It is clear that o must be even for the curve to close.

For instance, with 8 sticks you could form only the following regions: the square (of size 2 by 2, Fig. 4a), the triangle (Fig. 4b) and the rectangle (Fig. 4c).

The shapes of these regions are called discrete shapes.

<u>The shape number</u> of a discrete region (that is, of a region having a discrete shape) is obtained from that region by clockwisely replacing each salient corner by a 1, each straight corner by a 2, and each concave corner by a 3 (See Fig. 2g). Moreover, in order to obtain a



a 12121212 b 11212113 c 11221122

FIGURE 4 'ALL THE DISCRETE SHAPES OF ORDER 8'

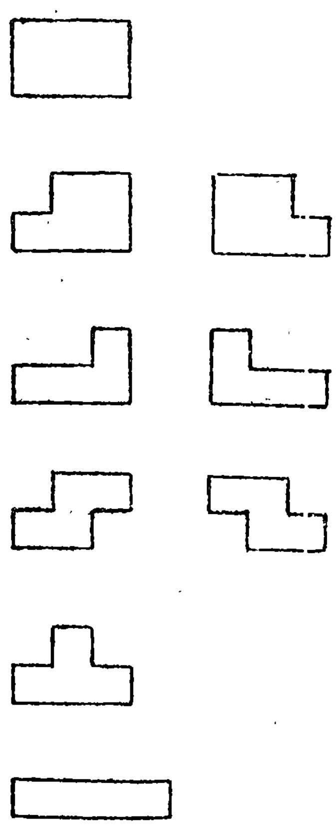


FIGURE 5 'ALL THE DISCRETE SHAPES OR ORDER 10' unique shape number, we start the procedure in the corner that produces a string (of 1, 2 and 3's) of minimum value.

For instance, the shape number of figure 4b is 11212113 , which was obtained by starting in the upper central salient corner and travelling clockwise (first right and then down). Had we started in the lower left corner, we would have obtained 11311212 which is rejected because its value (as a ternary number) is larger than 11212113.

The shape number of a discrete shape does not depend on a grid of fixed orientation or size; it can be derived directly from the region. It differs in this manner from "derivative of Freeman chain."

The shape number of a discrete shape is unique. It does not depend on its position, size or orientation.

The order of a shape number is the number of ternary digits it has. It is therefore equal to the number of corners (of types 1, 2 and 3 in Fig. 2g) that the discrete shape has.

It is also the number of sticks (segments of equal length) present in the discrete shape. It is always even. It is equal to the perimeter of the region.

All the discrete shapes of order 4

There is only one discrete shape of order four, the square. Its shape number is 1111.

This is the most primitive or fundamental shape. Imagine you are looking at things very far away; you can not really differentiate much. All objects would look round (square, in this paper) and equal.

All the discrete shapes of order 6, 8, 10 and 12

There is only one discrete shape of order 6, the rectangle with shape number 112112.

The three discrete shapes of order 8 are given in Fig. 4. Here the triangle appears for the first time.

The nine discrete shapes of order 10 are given in Fig. 5; those of order twelve are in Fig. 6. They are 36.

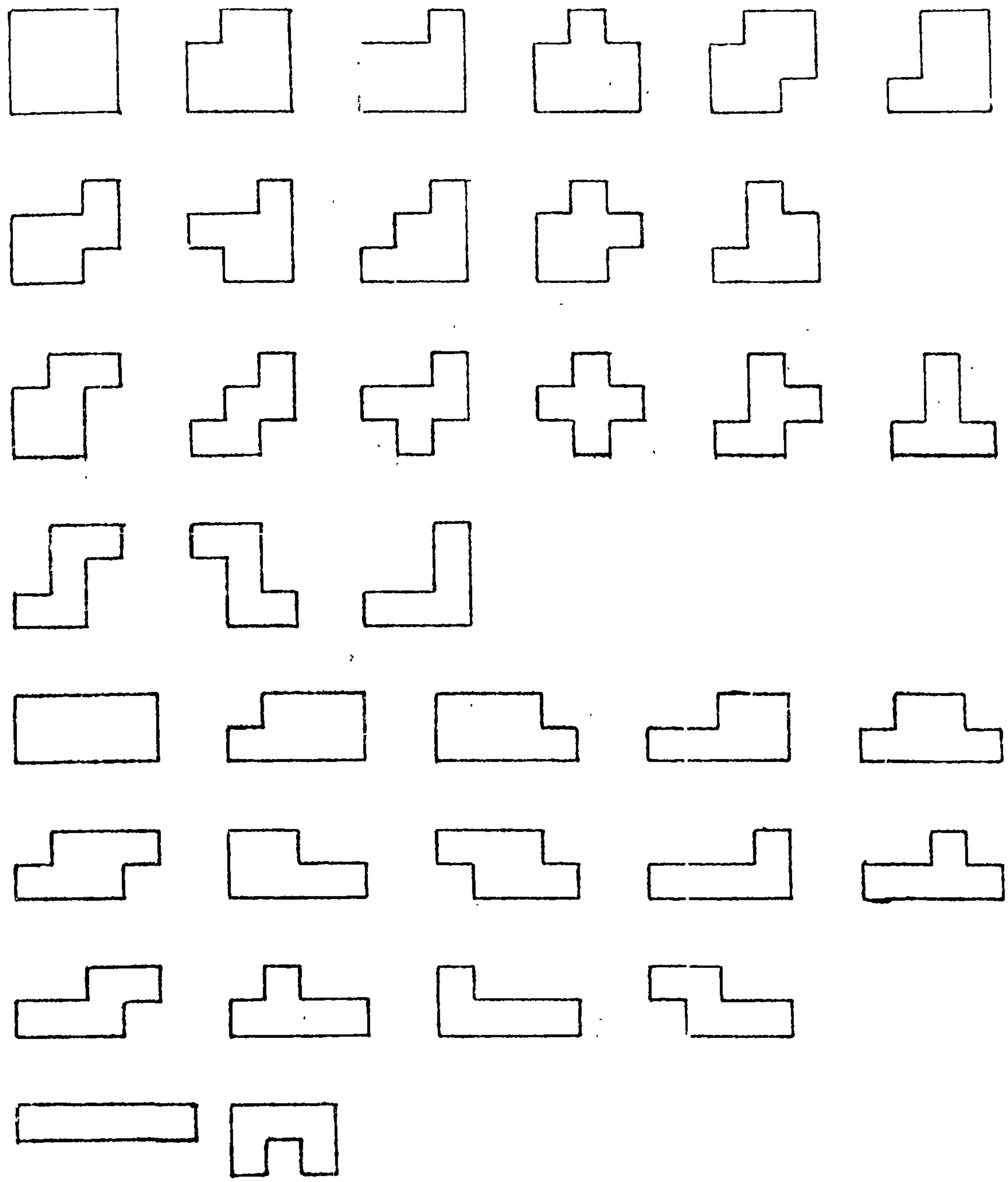


FIGURE 6 'ALL THE DISCRETE SHAPES OF ORDER 12'
The order of a shape indicates the number of sticks that are used to form it.

The discrete shape of a region

In order to find out the shape number of order o for an arbitrary region (and not just for the regions having discrete shapes), it is now only necessary to associate in some manner to that region a discrete shape, and then to give the shape number of that discrete shape to the region itself.

One way to proceed would be to compare (for instance, the areas in the least squares sense) that region with every discrete shape of order o (retrieved from a table such as Fig. 5) [6], and to select the discrete shape having the best fit (smallest error, best correlation).

Other way is given below, preferred because it does not use table lookup, back-tracking, error computation or pattern matching in the CONVERT [4] sense: we do not need to find out what is the distance or error between 11212113 and 12121212 , for instance.

<u>To find the shape number of order o of a region</u>:

1. Find out the basic rectangle and the excentricity of the region.
2. Select the rectangle with shape number o and excentricity closest to e. Align and center this rectangle over the basic rectangle of the region, thus defining a grid over the region.

   The orientation of the grid follows the basic rectangle, and the size of the grid is such that (a) every cell of the grid is a square, and (b) the basic rectangle has a shape number of order o for such grid. Already positioned, the rectangle selected in this step closely coincides with the basic rectangle.

   In practice, we have found better not to approximate the excentricity, but the sides of the rectangle instead. That is, select a rectangle with long side closest to $y=(o/2)(e/1+e)$.
3. Mark with a 1 each cell of the grid of step 2 that is more than 50% contained in the region.

   The collection of grid squares containning a 1 forms a discrete shape.
4. Find the shape number of the discrete shape of step 3, and give that as answer (but see discussion below).

An example is given in Fig. 7 'FINDING THE SHAPE NUMBER .'

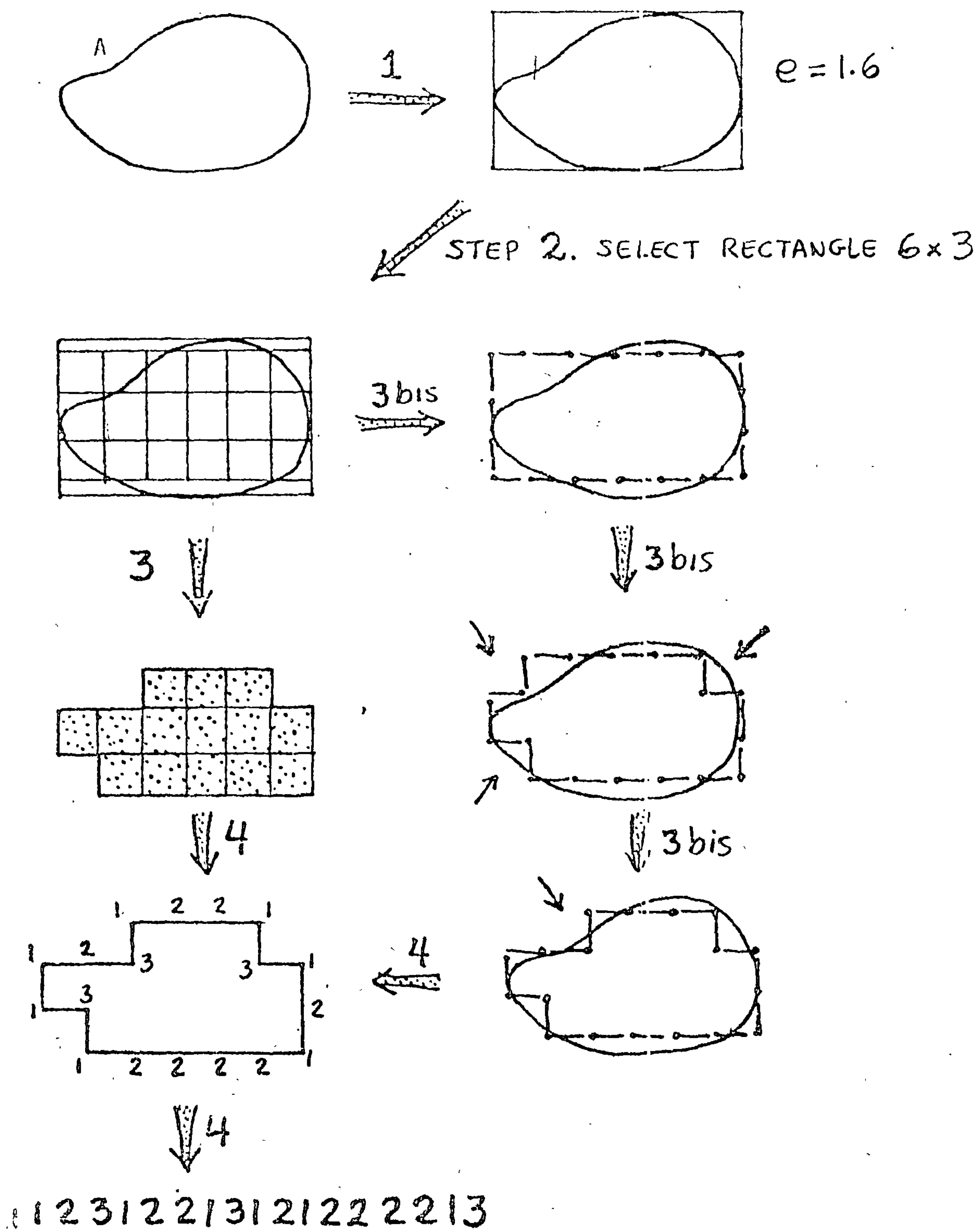


FIGURE 7 'FINDING THE SHAPE NUMBER'

The shape number of order 18 of region A is desired. The answer is 112312213121222213. The main procedure is through steps 1,2,3,4. Step 3bis is a (long) step that can be taken instead of step 3. See text.

Each figure carries its own shape number "within it."

Is the shape number found in step 4 indeed of order o? The crucial step is 3 above. The answer is discussed after an alternative step 3.

3bis (variant). On the perimeter of the square selected in step 2, place o sticks on the "wires" of the grid. Looking at each corner (of type 1 in Fig. 2g) of these sticks, push it and make it become a corner of type 3 if it surrounds a cell of the grill filled less than 50% with the region. Keep pushing corners (see Fig. 7, steps 3bis) until no further progress is possible. (Then go to step 4 above). This step 3b could be taken instead of step 3.

It is clear that this step does not alter the order of the shape number, since the number of sticks does not change.

What could increase the number of sticks (the length of the perimeter) is a depression in the boundary, because (Fig. 8) in order to sink stick a to position b we need two extra sticks. In this case we end up with a shape number of order o+2, or in general of order o + 2d, where d is the depth of the depression.

The way to correct this anomaly is to begin step 2 by selecting a rectangle not of order o (because we have just found that o produces a shape number of order o + 2d) but of order o - 2d, and then the depression will add 2d sticks to it, obtaining a shape number of the correct order.

Since a depression changes depth as the size of the grid varies, we may have to try step 2 with rectangles of order o-2d, o-2d+2, ..., o-2, until we find the shape number of order o.

Informally speaking, the order of the shape number is the degree of resolution being used to encode the shape.

The excentricity of the shape is important. It is a shape parameter coarser than the shape number. Two shapes of order o with basic rectangles of different excentricities can not be equal. The basic rectangle and the excentricity can be directly computed from the shape numbers (suggestion 6).

<u>Degenerate regions</u>. If the grid size is too large for some parts of a region, there will be totally blank squares that break the discrete shape into two or more pieces. Then the shape number of that region does not exist for that order. This is not an anomaly, but it is giving information regarding the minimum size grid for which a shape number makes sense (Fig. 8-II).

Meaningful shape order. A region with a very ragged and twisted perimeter will "demand" a higher order for a proper description than a region with smooth boundary; it expects more accuracy, because of its higher information content. [1] measures this appropriateness, also related to degenerate regions.

THE DEGREE OF SIMILARITY BETWEEN THE SHAPES OF TWO REGIONS

The shape number of a region enables us to find out instances of a given shape, even when distorted by enlargement or rotation. It answers the question "Have these two regions the same shape?", up to an order o.

In practice, however, a shape rarely repeats itself, due to noise and the allowable variations (for instance, ten sillouettes of apples have similar but not identical shapes). The relevant questions to answer are "How much different are these two forms?", "How much do these two shapes resemble each other?", "Is region A closer in shape to B, or to C?". This section gives a procedure to quantitatively answer these questions.

When the shapes of two regions A and B are compared, we can notice that the shape of order 4 of A, $s_4(a)$, is equal to 1111 (the only shape of order 4), and is therefore equal to $s_4(b)$.

Also $s_6(a) = s_6(b)$; probably $s_8(a) = s_8(b)$. It is likely that their first few shape numbers be identical. The reason is that the discrete shapes are coarse and not varied at low orders, where the "resolution" is low.

Nevertheless, most likely $s_{100}(a) \neq s_{100}(b)$, also $s_{98}(a) \neq s_{98}(b)$, etc. This is expected, because, due to the finer precision at higher orders, there exists a large variety of shapes, thus the discrimination between A and B is more demanding.

Of course, if A and B were very similar (but not identical), one could need to go up to say 170 to find that $s_{170}(a) \neq s_{170}(b)$. On the other hand, if they are visibly different (not alike at all), already at order 10 we will be having $s_{10}(a) \neq s_{10}(b)$.

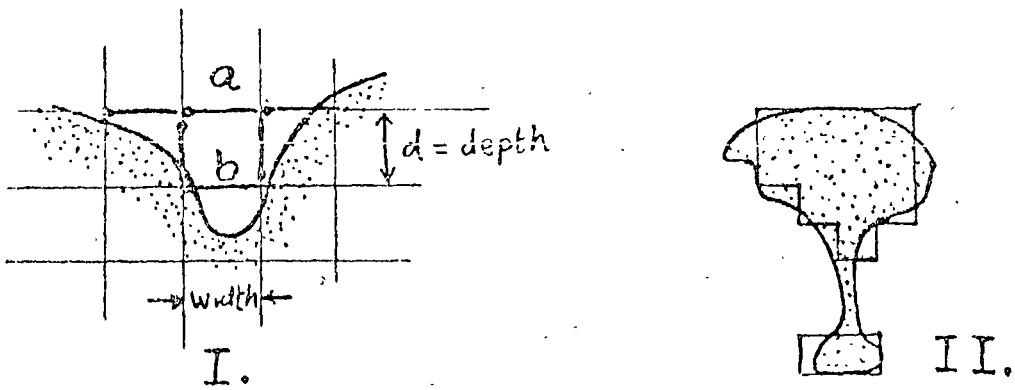


FIGURE 8 HOLES AND DEGENERATE SHAPES

I: A depression of depth d increases the shape number by 2d.

II: Degenerate regions split the discrete shape but do not have a shape number at this order.

Thus, as we increase the order o of the two shape numbers $s_o(a)$ and $s_o(b)$, they begin being equal but at some order they become different from that point on. How deeply they remain equal gives us an idea of the similarity between the shapes of a and b.

<u>Degree of similarity</u> k between the shapes of two regions a and b: It is the largest order for which their shape numbers still coincide.

That is, it is the largest m for which $s_m(a) = s_m(b)$, but $s_{m+i}(a) \neq s_{m+i}(b)$ for all i greater than 0.

That is, we have $s_4(a)=s_4(b)$, $s_6(a)=s_6(b)$, $s_8(a)=s_8(b)$, ..., $s_k(a)=s_k(b)$, $s_{k+2}(a)\neq s_{k+2}(b)$, $s_{k+4}(a)\neq s_{k+4}(b)$, ...

If a and b are regions with degree k of similarity, we write $a \simeq_k b$.

Example. For the figures of Fig. 9 'DEGREE OF SIMILARITY' we have for figures a, b and c:

$s_4(a)$= 1 1 1 1	$s_4(b)$= 1 1 1 1	$s_4(c)$= 1 1 1 1
$s_6(a)$= 1 1 2 1 1 2	$s_6(b)$= 1 1 2 1 1 2	$s_6(c)$= 1 1 2 1 1 2
$s_8(a)$= 11221122	$s_8(b)$= 12121212	$s_8(c)$= 12121212
$s_{10}(a)$= 1122211222	$s_{10}(b)$= 1131212122	$s_{10}(c)$= 1212212122
$s_{12}(a)$= 112221131213	$s_{12}(b)$= 121221221213	$s_{12}(c)$= 121222121222
$s_{14}(a)$=11222211231132	$s_{14}(b)$=12121312212123	$s_{14}(c)$=11312212212213

Therefore, a and b have a degree of similarity equal to 6: $a \simeq_6 b$.
a and c have a degree of similarity equal to 6, written $a \simeq_6 c$,
b and c have a degree of similarity equal to 8, written $b \simeq_8 c$.
This is represented both as a similarity tree (Fig. 9b) and as a similarity matrix (Fig. 9c) where other regions were also included.

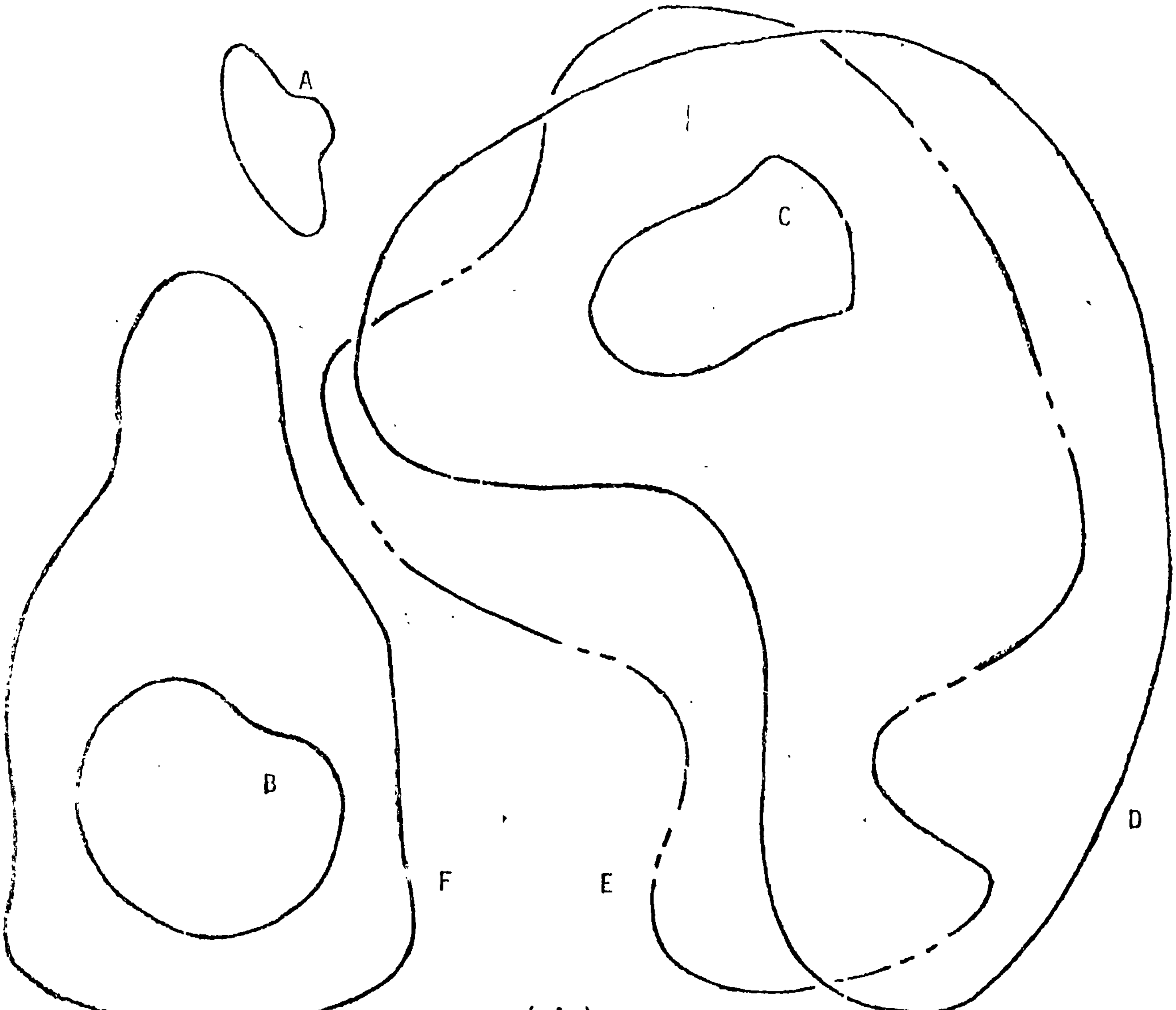
The similarity matrix is symmetrical; in fact, it is easily proved that, for arbitrary regions a and b,

(1) (Thm.) The relation "a and b have degree k of similarity" (for a fixed k) is <u>not</u> an equivalence relation, but

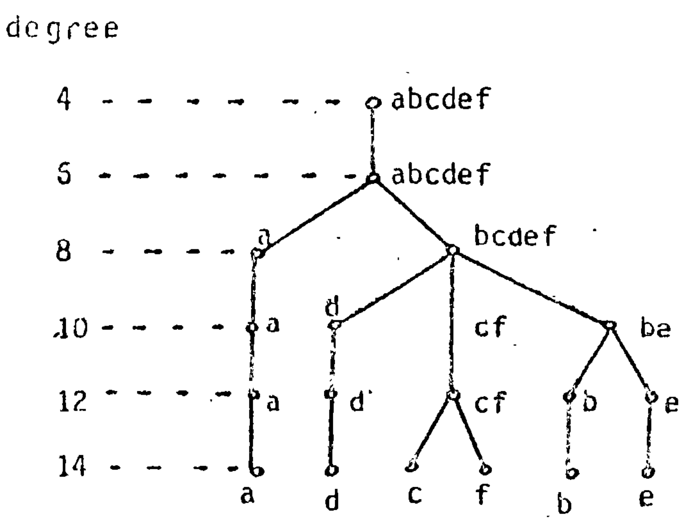
(2) (Thm.) The relation "a and b have degree of similarity of at least k" (for a fixed k) <u>is</u> an equivalence relation.

In fact, the equivalence classes of (2) for k=10 are nine and a canonical shape for each of them is given in figure 5.

Informally speaking, the size (power) of the magnifying lens that barely confuses two regions gives the degree of similarity between such regions.



(A)



(B)

	A	B	C	D	E	F
A	∞	6	6	6	6	6
B	6	∞	8	8	10	8
C	6	8	∞	8	8	12
D	6	8	8	∞	8	8
E	6	10	8	8	∞	8
F	6	8	12	8	8	∞

(C)

FIGURE 9 'DEGREE OF SIMILARITY'

(A) regions to be analyzed. (B) Similarity tree for (A).
(C) Similarity matrix for regions (A).

We could see the whole procedure as follows: A number is associated to each one of two regions. If the numbers are equal, the regions have identical shape. If not, another pair of numbers is deduced, and so on until we find that the two numbers coincide. The number of stages needed is an indication of the dissemblance of the two shapes.

Remarks on the degree of similarity

No parsing is necessary. To find the degree of similarity between a and b, shape numbers are compared for equality. Two shape numbers of different orders are inconmensurable (can not be compared, should not, need not).

Two shape numbers of the same order are either equal or different. If different, there is no need to compare "how close they are."

To find out the degree of similarity, a binary search is used: Is $s_8(a)$ equal to $s_8(b)$? Then compare at order 100 (the highest). Then at the middle. Then at the middle of the remaining valid half. And so on. A modified binary search [5] is better.

Wheatstone Bridge. In this old instrument to measure the value of resistances, an amperimeter says whether a current is zero or not. But this amperimeter does not measure the resistance itself; it only says: "current is zero. Stop!" Then the value of the resistance is obtained by a formula that does not involve the current (because it is zero!). Naturally, it does not need to be a high precission amperimeter.

In our case; the degree of similarity is not given by the shape numbers comparison test. It is given by a process that uses the comparison test.

Temperature readings. If the degree of similarity between a and b is 14, and that between c and d is 28, you can conclude that c and d are closer to each other than a and b. But we can not conclude that c and d are "twice as close in shape" as a and b. This is like the temperature: a body at 80°C is not twice as hot as one at 40°C (if you do not believe it, convert them to °F, or to °K). But see suggestion 9.

Ultradistance. If we define the distance between two shapes a and b to be the inverse of their degree of similarity, then we could easily prove that this is not only a distance, but it is also an ultradistance: it obeys $d(a,c) \leq \sup(d(a,b), d(b,c))$ in addition to the less demanding condition $d(a,c) \leq d(a,b) + d(b,c)$.

Comments on this theory of shapes

Shape numbers are not invariant under (1) reflections (mirror images); (2) skewing, where the figure is distorted by changing the angle between x and y; (3) unequal expansion, that is $X'=\varsigma_1 x$, $Y'=\varsigma_2 y$, with $\varsigma_1 \neq \varsigma_2$. This transforms a circle into an ellipse.

These transformations (1)-(3) alter what could be considered the (intuitive) shape of a figure. At the end of the paper a "Theory B" of shapes is presented, where condition (3) is violated, and therefore all circles and ellipses, disregarding size, excentricity, orientation, have the same Bshape numbers.

Problems with this theory of shapes

1. Ocassional loop in the similarity tree. Due to noise or the 50% requirement for quantization, and at low orders, sometimes it is observed a transitory divergence and then convergence in the shapes of two regions, v. gr., $s_8(a)=s_8(b)$, $s_{10}(a)\neq s_{10}(b)$, $s_{12}(a)=s_{12}(b)$, $s_{14}(a)\neq s_{14}(b)$, $s_{16}(a)\neq s_{16}(b)$, ... I.e., they were already different at order 10, but they are again equal at order 12 (however, only to separate soon forever). This still gives a unique shape number for a region, but makes the definition of the degree of similarity less attractive, and the procedure to find it, unreliable.

Only loops of size 2 (such as the example given) have been found, infrequently.

A way to make these loops disappear is to ignore half of the orders, for instance those not divisible by four. Orders 4, 8, 12, 16, ... remain. All the loops of size 2 have vanished (suggestion 8b).

2. Non existent shape numbers. Shape number of order o may ocassionally not exist for a given figure, due for instance to symmetrical holes of the type of figure 8.I. This does not bother the similarity procedure, but it is a nuisance not to have that shape number. See also suggestion 8a.

3. Quantization of the excentricity. For an object of excentricity 1.6 (Fig. 7), what rectangle will be used as the basis for computing its shape number of order 12? Will we use the 3 by 3 square (e=1) or the 4 by 2 rectangle (e=2)? An error will be commited in any case. You have to take one or the other. There seems to be no way out of this. See suggestion 5.

We now present a theory that has none of these problems.

THEORY "B" FOR SHAPE DESCRIPTION AND SHAPE COMPARISON

To obtain this new theory, we will make some changes to the old one:
1. Force the excentricity of any region to be equal to one, by performing an anisotropic dilation of its axis, $X' = \varsigma_1 x$, $Y' = \varsigma_2 y$. Now a circle and an ellipse will have the same Bshape; the Bshape of a rectangle will coincide with that of an square. As far as the <u>discrete</u> shapes, the only discrete Bshapes that now exist are those obtained from squares.

2. Do not go into depressions (fig. 8.I) with width smaller than the size of the cell of the grid. This avoids degenerate shapes (cf. also 'Reasonable shape numbers' above). That is, if a region is "scratched" by thin lines (thiner than the size of the grid) that belong to the background, either ignore them (act as if they were not there) or else, if they can not be ignored, this Theory "B" says that the size of this grid is inappropriate to describe such region, and that its Bshape number at this order does not exist.

3. Let these depressions (Fig. 8.I) generate Bshape numbers having a number of ternary digits larger than the expected order. That is, do not correct the anomaly that these depressions cause. The perimeter of the Bshape does not tell anymore its order.

4. Elliminate the orders that are not powers of two. The only valid orders for Bshape numbers are 4, 8, 16, ... These numbers still indicate the number of sticks to place around the basic square (remember, now a rectangle is converted first into a square) of the region (refer to step 3b of Fig. 7).

The procedure is the following:

<u>To find the Bshape number of order $o = 2^n$</u> of a region:
1. Find out the basic rectangle of the region and convert it into a square.

Declare that the Bshape number does not exist if the region has parts (necks, straights) or depressions (channels) narrower than 2^{2-n} or 4/o.

2. Make a grid by dividing the side of the basic square into o/4 parts.
3. Mark with a 1 each cell of the grid of step 2 that is more than 50% contained in the region (step 3bis given above could also have been used instead of this step 3). The collection of grid squares containning a 1 forms a discrete Bshape.
4. Find the shape number of the discrete Bshape of step 3, and give that as answer, even if it has more than o ternary digits.

<u>The order of a Bshape number</u> is four times the number of parts into which the side of the basic square was divided. It is also the perimeter (measured by the number of sticks) of the basic square.

It is no longer the number of ternary digits of the Bshape number, nor the perimeter of the discrete Bshape.

<u>The degree of similarity</u> between the Bshapes of two regions is obtained as before. Definition unchanged.

<u>Downwards constructability</u>. Given the Bshape number of order o of a region, the Bshape number of order o/2 can be deduced from it, by regroupping appropriate sets of four neighboring cells into a cell for the lower order. Therefore, if two regions have the same Bshape number of order o, they will continue to have equal numbers of smaller order, until they cease to exist. This gets rid of problem 1 'ocassional loops in the similarity tree' of the former theory.

<u>Upwards existence</u>. If the Bshape number of order o of a region exists, the existence of numbers for higher orders is guaranteed: (1) the inexistence of channels or isthmus of the region thiner than 4/o implies the inexistance of those narrower than 4/(o+i), for i>0; and (2) wider depressions (wider than fig. 8.I) will produce valid parts of the Bshape number, although its number of digits may increase. This defeats problem 2 of the former theory, "non existent shape numbers."

Finally, problem 3 of the former theory "quantization of the excentricity" is not present in Theory "B" because all excentricities are now equal to 1.

Nevertheless, we like more the former theory.

<u>Disadvantage of Theory "B"</u>. Squeezing along one axis is now a valid (Bshape preserving) transformation. Thus, either your application does

not care for the excentricity or aspect ratio, or you carry it as another parameter,in addition to the Bshape number. I suppose you are going to be carrying other parameters of the region (length, orientation) anyway.

Also, more care needs to be exercised now when selecting the major and minor axis, to avoid noise perturbations (cf. suggestion 7).

Suggestions and recommendations for further work

1. Use other tessellations (triangles, hexagons) instead of the square grid. I would like to see the triangle and circle as primitive shapes at low orders.

2. Use eight directions for the sticks, not four. This will produce more shape numbers of a given order, thus making the tables of figures 4-6 larger. But this is safe because the deduction of the shape number does not involve table lookup or comparison with these canonical shapes.

3. Apply these theories to Scene Analysis of coloring books [3]; chromosomes; silouettes of industrial parts on a conveyor belt; hand printed digits and zip codes; automatic taxonomy of shapes of shoes, airplanes, insects (their outline); texture description where the pictures are binary.

4. Extend these theories to shapes with holes inside them.

5. (Refer to problem 3 of the first theory and to step 2 of the procedure to find the shape number) a) Distort slightly the basic rectangle of the region, together with the region, so as to have it coincide exactly with the rectangle chosen in step 2: the grid is now of rectangles that are almost squares. b) Select in step 2 the rectangle of order o that minimizes the discrepancy between the areas of the region and of the rectangle.

6. Write a procedure to find the excentricity from the shape number. Hint: find the basic rectangle.

7. A better method to encase the region into a box is needed. Noise could introduce errors in length and position. Use the methods in [1-3].

8. (Refer to problems 1 and 2 of first theory): a) Of course, given an order (30, say) it is possible to find the best shape number of order 30 that fits the region, by comparing (in the least squares sense) the region with all the shapes of order 30. In this way the existence of a shape

number for any order and any region could be guaranteed [1]. I suggest to look for a procedure that avoids many comparisons but still gives back the shape number of order 30. This undiscovered method could be slower, since it will be used only when the normal procedure fails [6].

b) In order to make the loops vanish, do not use all orders. Even more, space them non linearly: use only orders 4, 6, 10, 16, 24, ...

9. Apply these theories to clustering. Do you want to group 200 figures into 24 classes according to their shapes? Construct their similarity tree, and cut it at a level such that the number of nodes at that level is approx. 24. You could answer relative likeness questions such as: "Is the difference between a and d larger than the difference between e and f?" The answer could be: "Yes, because $a \simeq_{10} d$ and $e \simeq_{14} f$." e and f went together longer. They needed a stronger lens, of order 16, to separate them.

Acknowledgments

To the M.I.T. Artificial Intelligence Laboratory (Boston) and to the hospitality of Prof. Marvin Minsky. The programs were finished at the A.I. Lab during a summer visit.

To the Ecole Nationale Supérieure des Télécommunications (Paris) and to Prof. C. Gueguen. This paper was finished by A.G. at ENST during a winter stay.

To IIMAS-UNAM (México) and to CETENAL (México), where the work was carried on.

References

1. Bribiesca, E., and Guzmán, A. Shape numbers: a rotation to describe pure form and to measure resemblance and likeness. Technical Report PR-78-18, April 1978. Computer Science Dept., IIMAS, National Univ. of Mexico. This report does not have the 6000 word limit of this paper.

2. Freeman, H., and Shapira, R. Determining the minimum-area encasing rectangle for an arbitrary closed curve. Comm. ACM 18,7,409-413, 1975.

3. Guzmán, A. Analysis of curved line drawings using context and global information. In Machine Intelligence VI (D. Michie and B. Meltzer, eds.) University of Edinburgh Press. 1971.

4. Guzmán, A., and McIntosh, H.V. CONVERT. Comm ACM 9,8,604-15, 1966.

5. Nadler, M., remarked at his Seminar on Pattern Recognition (IRIA, France, Feb. 78) that since it costs more to compare larger orders than smaller orders, do not compare at the middle point, but move instead towards the extreme with the cheapest test by an amount proportional to the ratio of high to low costs.

6. This method could produce different shape numbers than those found by the procedure described in the paper. Both methods are not equivalent.

COMUNICACIONES

TECNICAS

1977

142. Ignacio Méndez y Germán Rivera. ON THE USE OF FERTILITY TREND FUNCTIONS FOR COMPARING A LARGE NUMBER OF VARIETIES.

143. Angel Kuri. UN MODELO PROBABILISTICO DE UN SISTEMA DE INFORMACION

144. Joaquín Díaz. ESTIMATION OF RELIABILITY FOR A GENERALIZED GAMMA DISTRIBUTION

145. Luis A. Haro y Arturo G. Hermosillo. CIRCUITO DE ENCENDIDO AL TOQUE Y TIMER

146. José Angel Canavati y Antonmaría Minzoni. A DISCONTINUOS STEKLOV PROBLEM WITH APPLICATION TO WATER WAVES. *

147. I. Herrera, A Minzoni y E. Flores. THEORY OF FLOW IN UNCONFINED AQUIFERS BY INTEGRODIFFERENTIAL EQUATIONS

148. Jean Pierre Hennnart. A FINITE ELEMENT MODEL FOR PLASMA SIMULATION IN MULTIPOLES

149. Federico O"Reilly y Joaquín Díaz. ON THE APPLICATIONS OF SEQUENTIAL UPDATING IN REGRESSION TO OBTAIN U (0,1) INDEPENDENT [illegible] AND SOME UNBALANCED DESIGNS

150. Charles Boyer y Jerzy Plebañski. GENERAL RELATIVITY AND G STRUCTURES. I. GENERAL THEORY AND ALGEBRAICALLY DEGENERATE SPACES.

151. Javier F. Aranda e Ignacio Méndez. ANALISIS DISCRIMINANTE

152. E. Villegas y M. Padilla. DISEÑO DE UN PROGRAMADOR PARA LOS PROM'S 1602,1602 A y LOS EPROM'S 1702,1702 A.

153. Arturo G. Hermosillo y Arturo Haro. DISEÑO Y CONSTRUCCION DE UN MODEM DE 1200 BAUDS

154. Gursharan S. Sidhu y Howard L. Weinart. VECTOR-VALUED Lg-SPLINES. [illegible] *

155. Roland England, Alejandro Chagoya, René González y J. Moisés Pelayo A MARKOVIAN MODEL FOR SIMULATION OF STUDENT FLOW AND FORECASTING OF ENROLLMENT IN MEXICAN PRIMARY SCHOOLS.

* Ver lista de FE DE ERRATAS DE COMUNICACIONES TECNICAS

156. Shri Krishna Singh y Federico J. Sabina. MAGNETIC ANOMALY DUE TO VERTICAL RIGHT CIRCULAR CYLINDER WITH ARBITRARY POLARIZATION.

157. Renato Barrera, Jennie Becerra, Ramón Castillo, Maximiliano Díaz, Víctor Guerra, Adolfo Guzmán, Patricia Huacuja, Armando Jinich, Juan Manuel López, Eduardo Llera, Juan Ludlow, Rosa Seco y Víctor Germán Sánchez. DETECCION Y CUANTIFICACION DE RECURSOS AGROPECUARIOS MEDIANTE ANALISIS POR COMPUTADORA DE FOTOGRAFIAS TOMADAS DESDE AVION Y SATELITE.

158. Cinna Lomnitz. MISION SISMOLOGICA A LA REPUBLICA POPULAR CHINA. 16 al 30 de Mayo de 1977. NOTAS DE CAMPO. ($)

159. José Luis Abreu CAN MATHEMATICS BE A BASIS FOR PHILOSOPHICAL KNOWLEDGE .

160. José Luis Abreu. ON INTUITIONISM AND CONSTRUCTIVISM

161. Joaquín Díaz. BAYESIAN ESTIMATION OF RELIABILITY FOR A GENERALIZED GAMMA DISTRIBUTION.

162. Antonio Aguilar y Kurt Bernardo Wolf. SYMMETRIES OF THE SECOND - DIFFERENCE MATRIX AND THE FINITE FOURIER TRANSFORM

163. Alejandro Velasco Levy. PROYECTO RAMSES. RED AUTOMATICA MICROMETEOROLOGICA Y SISTEMA DE ECO-SONDEO.

164. Alejandro Guarda A. MODELADO SE TRANSISTORES. (1a parte): MODELO DEL SEMICONDUCTOR.

165. Alejandro Guarda A. MODELADO DE TRANSISTORES (2a Parte): MODELO DE LA JUNTURA P-N.

FE DE ERRATAS DE COMUNICACIONES TECNICAS 1977

* 146. Angel Canavati y Max Díaz. ON APPROXIMATION NUMBERS, n-DIAMETERS AND MEASURE OF NON-COMPACTNESS.

Este trabajo está en elaboración. Por lo tanto su lugar lugar lo ocupa el actual 146.

---

* 154. Por error de clasificación de contenido este número lo ocupaba:

Manuel Padilla. DETECTOR DE LLAMADAS DE LARGA DISTANCIA

Dicho trabajo pertenece a la nueva Serie Amarilla: Desarrollo No. 5.

---

26. Ignacio Méndez. EXPERIMENTOS FACTORIALES CONFUNDIDOS

27. Fausto Ongay Larios. ESTUDIO MATEMATICO DE LA ECUACION DE HARTREE

28. Ignacio Méndez. EXPERIMENTOS FACTORIALES FRACCIONALES

29. Ignacio Méndez. CONSIDERACIONES RELATIVAS A LA POSTULACION Y CONTRASTACION DE HIPOTESIS CIENTIFICAS.

30. Ignacio Méndez. METODOLOGIA DE SUPERFICIE DE RESPUESTA

31. Ignacio Méndez. MODELOS MIXTOS Y ALEATORIOS EN EL DISEÑO Y ANALISIS DE EXPERIMENTOS.

1977———————————Vol.1———————————Serie Verde:Notas

6. J.M. Fernández Peña. MICRO SDJ8 MANUAL DE USUARIO.

7. Carlos Ruiz de Velasco. SIMCIRDG. SIMULADOR DE CIRCUITOS DIGITALES

8. Jorge Stephenson. TRANSMISION DE DATOS Y MODEMS

9. Guillermo Espinosa y Arturo López. INTRODUCCION A LOS METODOS JERARQUICOS DE ANALISIS DE CUMULOS

1978--------------------Vol.2--------------Serie Verde:Notas

10. Jorge Ize. LAS ECUACIONES EN DERIVADAS PARCIALES Y SUS APLICACIONES (CURSO DE VERANO) TEORIA DE EXISTENCIA PARA ECUACIONES EN DERIVADAS PARCIALES.

11. Alejandro Guarda A. INTRODUCCION AL PROCESAMIENTO DIGITAL

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

SISTEMAS AVANZADOS DE GEO-INFORMATICA EN PLANEACION

PROGRAMAS DE GEO-INFORMATICA
INTERCAMBIO DE PROGRAMAS GEOGRAFICOS
UNIVERSIDAD ESTATAL DE MICHIGAN

JULIO, 1978.

Palacio de Minería Calle de Tacuba 5, primer piso. México 1, D. F. Tels: 521-40-23 521-73-35 5123-123

THREE NEW COMPUTER PROGRAMS FOR SPATIAL ANALYSIS AND COMPUTER MAPPING

The Computer Institute for Social Science Research is now distributing three new programs in its spatial analysis series. These programs are (1) GEOSYS (Version 4): A Computer System for the Description and Analysis of Spatial Point Data; (2) FLOW (Version 4): A Computer System for the Analysis and Graphic Description of Spatial Networks; and (3) CASP: Choropleth Area Shading Program. The GEOSYS system is described in an article by R. I. Wittick appearing in the July, 1976, issue of Geographical Analysis, and the FLOW system is described in an article by R. I. Wittick appearing in the May, 1976, issue of the Southeastern Geographer.

All three programs are written entirely in FORTRAN for a Control Data 6500 computer. CASP has approximately 1,000 source statements, FLOW is overlayed and has approximately 4,000 statements, and GEOSYS is overlayed with approximately 6,500 source statements. Abstracts for the three programs are given on the following pages.

The user documentation for all three programs may be purchased from the Computer Institute for Social Science Research for $1.00. Please send your check or money order (do not send cash) for $1.00 made payable to Michigan State University to:

Ms. Kristen Halle
505 Computer Center
Michigan State University
East Lansing, MI 48824

The programs are distributed on magnetic tape (either 7 or 9 track - please specify when ordering), and the prices given below include a new

2400 foot magnetic tape. The following is sent with each order:

1. Three copies of the printed user documentation
2. One magnetic tape containing two files:
   File 1 - source code and test data
   File 2 - test output
3. Copies of any test plots generated by the test data

The programs are priced as follows:

CASP	$ 75.00
FLOW	125.00
GEOSYS	195.00

When ordering one or more of these programs, please send a check, money order, or purchase order for the total amount of the order. Checks and money orders should be made payable to Michigan State University and sent to:

Ms. Kristen Halle
505 Computer Center
Michigan State University
East Lansing, MI 48824

FLOW: A Geographic Computer System for the Analysis and Graphic Depiction of Spatial Flow Data

FLOW is a large overlay system designed to analyze and map qualitative or quantitative flow data. The mapping modules allow for the depiction of the linkages, with or without volumes assigned to them, the depiction of the nodes defined on either interval or nominal scales or both, and the depiction of one or more area outlines. Options are also included for map title, border, scale, legends, and plotting characteristics. The system can be run under batch or interactive modes, and the interactive user can view the maps on a Tektronix graphics terminal if desired; otherwise, all plotting is routed to a pen-and-ink plotter.

The system includes analysis modules which compute graph theoretic measures, node accessibilities, shortest paths through the network, two gravity models, and a transportation model. A free form input structure is used for inputting commands and data.

The system is written entirely in FORTRAN and organized into overlays. Core storage is dynamically allocated according to the size of the data set(s) being used, thus minimizing the amount of central memory needed.

The network structure is defined to FLOW by (1) Cartesian coordinates of the nodes, (2) optional values for node sizes, (3) node-pair identifications for the linkages, (4) optional values for link volumes, and (5) optional nominal categories for the nodes. In addition, unidirectional flows may be designated if desired.

PROGRAMMING BASIS/OPERATING INFORMATION:

Mode of usage: Batch and/or interactive (including interactive graphics)

Operative computer: CDC 6500

Memory requirements: Loads in 26,000 octal words; allocates memory dynamically.

Peripherals: Pen-and-ink plotter (optional); Tektronix graphics terminal (optional); two tape or disk files for optional data input.

Programming language: FORTRAN (with overlays)

Operating system: SCOPE; however, should run on any system with overlay capabilities

Data files used: (1) Node identifications file, (2) link identifications file, and (3) area outline coordinate file

Approximate number of source statements: 4000

GEOSYS: A Geographic Computer System for the Analysis and Graphic Depiction of Spatial Point Data

This system consists of several components or modules, each of which performs a particular set of analytic or graphic functions. GEOSYS can be accessed by submitting a deck of punched cards to the computer or by typing commands at a remote terminal. If a graphic terminal is used (such as a Tektronix 4000 series), the map output can be displayed on the terminal screen; otherwise it is sent to a pen-and-ink plotter.

The current operating version of GEOSYS includes eight analysis modules: (1) centrographic analysis, (2) ring and sector counting, (3) spatial interpolation, (4) nearest neighbor analysis, (5) binomial weighted smoothing, (6) trend surface analysis, (7) a module for evaluating trend surface equations to geographic matrices, and (8) a module for the determination of geographical origins. The present graphic capabilities of the system include both a line printer and a plotter contour mapping package, a plotter block diagram module, a line printer and a plotter point symbolic mapping module, and a plotter module for portraying three dimensional histograms. Four miscellaneous modules are also provided for (1) computing spherical distances from latitude-longitude coordinates, (2) creating input data to the SYMAP program, (3) extracting coordinates from distance matrices, and (4) converting latitude-longitude coordinates to Cartesian coordinates.

When using the system, a module is executed by a single command word punched on a card or typed at a terminal. Almost all of the module commands have optional parameters which can be added to provide greater flexibility in the use of the module. Default values are included for all parameters, however as an aid to the user who is unfamiliar with the system or a particular module. In addition, extensive and clearly written diagnostics are provided to help the user in running the system.

Both coordinate data and geographical matrices (spatial lattices or gridded data) can be used with GEOSYS. Coordinates can be input with or without data or "Z" values. For modules such as the nearest neighbor analysis, the data values, if present, are ignored. Other modules, the trend surface analysis for example, require data values, and their omission will cause an error diagnostic to be generated. The use of data values for the centrographic analysis module and the ring and sector count module is optional; however, the results generated will vary depending on whether or not data values are included. That is, these analyses can be run with either weighted or unweighted point sets, and the results must be interpreted accordingly. For example, the mean center statistic of the centrographic analysis module should be interpreted as a geographic center when used with unweighted data points, but as a weighted mean center when data values are included.

Three different types of coordinate data can be used with GEOSYS. The first and most common consists of standard Cartesian coordinates. The second type allow for the special format required in the SYMAP program (i.e., the B-DATA POINTS and E-VALUES packages). The third type consists of latitude-longitude coordinates.

GEOSYS also accepts outline coordinate strings defining the perimeter or geographic extent of the area under study. The outline coordinates can be defined in the same manner as the data points; i.e., standard Cartesian coordinates, latitude-longitude coordinates, or SYMAP formatted coordinates (A-OUTLINE package).

The system is structured such that any number of modules can be executed in a single run provided the required type of data has been either input or computed by a previously executed module. Thus analysis chaining is possible whereby the output from one module serves as the input to a subsequently requested module. All necessary linkages for this type of operation are handled internally by GEOSYS, and are therefore transparent to the user. If a module is requested for which the necessary type of data is not available, an appropriate diagnostic will be included in the user's printed output or typed at his terminal if he is accessing GEOSYS interactively.

GEOSYS is written entirely in FORTRAN so that it can easily be implemented at other installations. The modular construction allows for simplified addition of new routines to the system. GEOSYS is based on an overlay structure. Therefore, even though it consists of over six thousand FORTRAN statements, only small portions of this code are in the computer's core storage at any one time. It should thus be possible to implement GEOSYS on computers with rather limited core storage. In addition, actual data storage is dynamically allocated. As a result, the only constraint on the amount of data being manipulated by GEOSYS is the available core storage of the host computer.

PROGRAMMING BASIS/OPERATING INFORMATION

Mode of usage: Batch and/or interactive (including interactive graphics).

Operative computer: CDC 6500

Memory requirements: Loads in 25,000 octal words, allocates memory dynamically.

Peripherals: Pen-and-ink plotter (optional); Tektronix graphics terminal (optional); four tape or disk files for intermediate storage and optional data input.

Programming language: FORTRAN (with overlays).

Operating system: SCOPE; however, should run on any system with overlay capabilities.

Data files used: any one or more of the following: (1) X-Y-Z data, (2) SYMAP formatted data, (3) gridded data, (4) latitude-longitude data, (5) area outlines in Cartesian, SYMAP, or latitude-longitude coordinates.

Approximate number of source statements: 6500

CASP: Choropleth Area Shading Program

CASP allows the user to simultaneously depict choropleth areas on two different scales: a nominal scale and an interval scale. The nominal scale is portrayed by the angle of the shading vectors, and the interval scale is depicted by the interval or gradient between the vectors. This capability is useful in the portrayal of such phenomena as land uses where there is an interest in both the types of land use and the intensity of use. The various land use types are depicted by the vector angles while the land use intensities are portrayed by the vector gradients.

Normally the use of two scales is not needed in choropleth mapping. Therefore, CASP also has the capability of depicting choropleth areas on a single scale, either nominal or interval. In addition, the variable to be mapped may be shown by either the angularity of the shading vectors or the gradient.

The data requirement for CASP are minimal. The user defines each of the choropleth areas (or polygons) in terms of those Cartesian coordinates which form the vertices of the polygon. The user must also define the data value or values for each area, and any of the program options he wishes to utilize. These options include the definition of map scale, title, border, legend, class intervals, and the location of the coordinate file (punched cards, magnetic tape or disk). All options have default values, and thus none are required if the user chooses to omit them. Once the map has been drawn the program will look for additional instructions, at which time the user may make option changes from the previous map (new class intervals or change of scale for example), add new data values, or submit an entirely new map. If additional instructions are not found, the program terminates.

Although the cost of each map varies with scale, number of choropleth area, and shading gradient, very few maps produced by the program cost over $2.00 for computer time and $3.50 for plotter time. These costs will naturally vary with the computer environment used and the particular charging rates employed. The program is written entirely in FORTRAN and uses only the standard plotter subroutines: PLOT, SYMBOL, and NUMBER. Thus, it can easily be adapted to computer models other than the CDC 6500 for which it was first written.

PROGRAMMING BASIS/OPERATING INFORMATION

Mode of usage: batch

Operative computer: CDC 6500

Fast Memory and Peripherals Required: dynamically allocates storage; pen-and-ink plotter, optional tape or disk for data input

Programming language: FORTRAN

Operating system required: SCOPE; however, should be independent of operating system

Data files used: file containing polygon definitions, and cards containing data values

Number of source statements: 1000

GEOGRAPHY PROGRAM EXCHANGE

The Geography Program Exchange (GPE) has been established to assist universities and other non-profit organizations with the interchange of computer software which relates to problems of a geographic nature. The GPE operates on a self-sustaining basis. Its services are made available to users anywhere in the world.

The GPE maintains a central file of specialized computer programs with associated documentation and test data sets, and makes copies of all or part of this material available to users at cost. The programs held by the GPE represent specialized materials developed for use in geographic research and teaching; no attempt is made to duplicate standard statistical programs. Distribution of available programs is in the form of listings, punched cards, or magnetic tapes.

Program holdings consist of items from the specialized program libraries of a large number of geography departments in North America as well as programs submitted by geographers from other countries. The GPE actively solicits programs from all sources. If the program submitted appears to be of enough general interest, it is added to those available for distribution. All software handled by the GPE is obtained and distributed on a non-royalty basis.

Although the GPE operates with limited resources, it does attempt to provide additional assistance to those departments and individuals who are just starting to make use of the digital computer as a research and teaching tool.

Overall management of the GPE is provided by a board consisting of Professors Duane F. Marble (Suny/Buffalo), Waldo R. Tobler (University of California-Santa Barbara), and Robert I. Wittick (Michigan State University). Dr. Wittick is in charge of operations and all inquiries relating to activities of the GPE should be directed to him at the following address:

Dr. Robert I. Wittick
Computer Institute for Social Science Research
Michigan State University
East Lansing, Michigan 48824
U.S.A.

Telephone: 517/353-2040

A European outlet for the GPE programs has been established in West Germany. For further information on the availability of programs from West Germany, write to:

Dr. Andre Kilchenmann
Geographisches Institut
Universitat Karlsruhe (TH)
75 Karlsruhe 1, Kaiserstrasse 12
Germany, BRD

INDEX OF PROGRAMS CURRENTLY AVAILABLE

June 1977

ACCESS: Tape Library Utility Program

A utility program which is used to create and edit a tape library of computer programs. Programs on tape may be listed, punched, stored on another tape, or compiled.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Contributor: Department of Geography
Northwestern University

ADJUST: Matrix Adjustment

A program designed to perform iterative multiplicative adjustment of matrices to desired marginal totals. The program can also be used in the analysis of transaction flows.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: July, 1972
Similar Programs: INDIFF
Contributor: Department of Geography
Michigan State University

ALLOC: Heuristic Solutions to Multi-Facility Location Problems on a graph

Program to compute optimal locations for facilities based on minimized aggregate distance using Marazana, and Tietz and Bart Algorithms.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Contributor: Department of Geography
University of Iowa

ALTERN: Multi-Source Location-Allocation Algorithm (Heuristic)

This program alternately locates sources at the minimum point of their assigned sinks and allocates sinks to their closest sources to approximate the optimal location for M sources to serve N sinks in a Euclidean plane.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Similar Programs: LAP, MATRAN, TORN
Contributor: Department of Geography
University of Iowa

AZMAP: Azimuthal Map Transformation

The program calculates and produces azimuthal map transformations on the pen-and-ink plotter, as a means of displaying geographic data.

Computer: IBM 360/50
Compiler: FORTRAN G
Date Received: May, 1971
Contributor: Department of Geography
University of Rhode Island

Other Versions: (AZMAP)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: January, 1972
Contributor: Computer Institute for Social Science Research
Michigan State University

BASICS: Calculation of Basic Statistical Series

Program to calculate and plot (on pen-and-ink plotter) one, or all of the following, which are commonly used in basic analyses of statistical series: (1) a histogram, (2) a data array, and (3) a cumulative-frequency graph.

Computer: CDC 6600
Compiler: 6000 FORTRAN EXTENDED
Date Received: March, 1972
Contributor: Department of Geography
University of Minnesota

BIVAR: Bivariate Means and Medians

Computes bivariate means and approximates bivariate medians by iterative means to a specified accuracy or a prescribed number of iterations. Also computes geometric coordinates, and total travel to the means and medians.

Computer:	PDP-10
Compiler:	FORTRAN IV
Date Received:	March, 1972
Contributor:	Department of Geography Western Michigan University

BLCK: Block Diagram Plotting

The program is used for rapid pen-and-ink plotting of isometric profiles from a rectangular matrix of data

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	August, 1970
Similar Programs:	ISOMET; MAP 3D; PERS; SURGE; TRID
Contributor:	Department of Geography University of Michigan

Other Versions: (BLCK)

Computer:	CDC 6500
Compiler:	6000 FORTRAN EXTENDED
Date Received:	December, 1971
Contributor:	Computer Institute for Social Science Research Michigan State University

CANON: Canonical Correlation

The program calculates the interrelationships between two sets of measurements made on the same observations.

Computer:	IBM 360/65
Compiler:	FORTRAN G
Date Received:	May, 1970
Contributor:	Department of Geography University of Iowa

CANTRN: Canonical Trend-Surface Analysis

This program determines geographic patterns of several variates simultaneously, thereby producing a multivariate trend surface. The purpose is to determine linear combinations for both geographic coordinates and the variates, Z (J), so as to maximize correlation between the two linear expressions.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: DYAD, MPLSQ, POLYFIT, TREND, TRENDC
Contributor: Department of Geography
State University of New York, Albany

CART1: Irregular Cartograms

Computes cartograms from areas defined by irregular polygons.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: October, 1971
Similar Programs: CART2
Contributor: Department of Geography
University of Michigan

CART2: Regular Cartograms

Computes regular cartograms from areas defined by a geographical matrix

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: October, 1971
Similar Programs: CART1
Contributor: Department of Geography
University of Michigan

CENDA: Distance-to-Centroid Discriminant Analysis

Classifies subjects on the basis of distance to group centroids.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: November, 1972
Similar Programs: COSINE, DISITR, ITERIM, MDISC, NEARNBR, NOLIDA
Contributor: Department of Geography
University of Waterloo

CENTRO: Centrographic Measures

The program computes descriptive measures of spatial distributions from coordinate data. Weighted or non-weighted point sets are allowable. Output includes mean center, standard radius, coefficient of circularity, and angle of orientation of the distribution.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: May, 1970
Similar Programs: ELIPS
Contributor: Department of Geography
University of Iowa

Other Versions: (CENTRO)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: August, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

CHIINT: Interactions for Chi-Square

To compute interactions for chi-square

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Contributor: Department of Geography
Northwestern University

CHOROS: Application of Linear Neighbornood Operators to Choropleath Maps

Program to apply linear neighborhood operators to choropleath maps by using the reaggregation method devised by Tobler.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: October, 1972
Contributor: Department of Geography
University of Michigan

CLCOUNT: Point Counting Within Rectangular Cells of a Rectangular Region

To count the number of points within rectangular cells of a rectangular region.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Contributor: Department of Geography
Northwestern University

Other Versions: (CLCOUNT)

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Contributor: Department of Geography
University of California, Los Angeles

CLIMAT: Recognition and Classification of Climate Types

A set of CAI programs for analyzing, cataloguing, and classifying raw climate data using a logical, uniform, and open ended system which is easily capable of being plotted and mapped.

Computer: GE-635
Compiler: GE BASIC
Date Received: March, 1972
Contributor: Department of Geography
Dartmouth College

CLUSTER: Clustered Pattern Recognition

The program establishes the location and areal extent of point clusters in a spatial distribution.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: LINEAR, REGLAR
Contributor: Department of Geography
State University of New York, Albany

CLUSTR: Cluster Analysis

Provides a means of grouping variables on a set of observations using cluster analysis. The clustering of variables begins with the pair of variables having the highest correlation and then proceeds by adding the variables having the highest correlation with the elements of the cluster.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: January, 1972
Similar Programs: DON
Contributor: Computer Institute for Social Science Research
Michigan State University

CMAP: Choropleth Mapping

Designed to produce choropleth maps utilizing small computers.

Computer: IBM 360/40
Compiler: FORTRAN E
Date Received: December, 1971
Similar Programs: CPLETH, XMAPP, COBMAP, INTRMAP
Contributor: Department of Geography
University of Idaho

COBMAP: Choropleth Mapping in COBOL

This program is a COBOL version of CMAP. The program will produce one map for each submission.

Computer: IBM 360/65
Compiler: ANS COBOL
Date Received: March 1976
Similar Programs: CMAP, CPLETH, XMAPP, INTRMAP
Contributor: Department of Geography
University of Tennessee

CNGRP: Contiguity Grouping

The program groups observations using the criterion of euclidean proximity in a p-dimensional vector space if the data are given in the form of geographical matrices, the program can automatically impose a geographical contiguity constraint on the grouping.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Similar Programs: CONGRP, DISGRP
Contributor: Department of Geography
University of Michigan

Other Versions: (CNGRP)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: August, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

CNTOUR: Contouring a Grid

A package of FORTRAN subroutines to produce a contour map (pen-and-ink plot) of a set of data points which form a grid. The grid does not necessarily have to be regular.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1972
Similar Programs: CONTR, CONTUR
Contributor: Computing Centre
University of British Columbia

COLMOG: D-Statistic for Kolmogorov-Smirnov Test

The program computes the d-statistic for the Kolmogorov-Smirnov Test for equivalent rows or columns of the two matrices output from CONDIST.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: June, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

CONDIST: Relationship of a Population to Distributed Facilities

The program computes the spatial relationship between a sample of points and a distribution of different sized facilities. The program computes the number of points within specified sizes. Also computed are the distances separating successive percentile groups of the points from their nearest facility larger than a specified size.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: June, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

CONGRP: Contiguity Grouping

Stepwise multivariate grouping of observations using one of several distance criteria. Contiguity restraints may be imposed upon the groupings.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: CNGRP, DISGRP
Contributor: Department of Geography
Northwestern University

Other Versions: (CONGRP)

Computer: IBM 360/65
Compiler: FORTRAN G
OS/360 Assembler-F
Date Received: May, 1970
Contributor: Department of Geography
University of Iowa

CONRAT: Contiguity Ratios

Computation of statistics for contiguity measures to provide an evaluation of the hypothesis of randomness in the areal arrangement of the values of a specified property. The program is used for a two-category nominal, or binary, scaling of values.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: KCOLOR, VALRAT, VALRATI
Contributor: Department of Geography
Northwestern University

CONTOUR: Contour Mapping

Construction of a contour map on the line printer from a series of data points whose locations are defined in terms of grid squares.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: LATMAP, RGRID
Contributor: Department of Geography
Northwestern University

CONTR: Automatic Contour Mapping

The program produces contour maps either on a pen-and-ink plotter, or on a cathode ray tube. The program will accept either a rectangular or an irregular shaped (not spaced) data set within a rectangular area.

Computer: IBM 360/65
Compiler: FORTRAN C
Date Received: May, 1970
Similar Programs: CNTOUR, CONTUR
Contributor: Department of Geography
University of Iowa

CONTUR: Contour Mapping

The program uses a pen-and-ink plotter to draw contour maps from data given in the form of geographical matrices. Stereograms and perspective contours can also be obtained. An option allows conversion of the contours to a map projection before plotting.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Similar Programs: CNTOUR, CONTR
Contributor: Department of Geography
University of Michigan

CONWGT: ontiguity-Structured Class Limits

The program weighs data values at each enumeration area by considering the spatial arrangement and similarity of the data units.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

COORD: Coordinate Conversion

Program to convert section, township, and range notation to Cartesian coordinates.

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Contributor: Department of Geography
University of California, Los Angeles

CORD: Generation of Coordinate Information for Mapping Routines

Coordinate systems are made readily available for use in the SYMAP, GIPSY, and Trend Surface mapping routines.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

COSINE: Correlation Discriminant Analysis

Discriminant analysis based on angular proximity to group centroids.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: November, 1972
Similar Programs: CENDS, DISITR, ITERIM, MDISC, NEARNBR, NOLIDA
Contributor: Department of Geography
University of Waterloo

COVAR: Covariance Analysis

Performs a covariance analysis on a set of multiple regressions. An unequal N size is allowed for the regressions. Includes tests for homoscedasticity, equality of regressions, and equality of origins.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: May, 1970
Contributor: Department of Geography
University of Iowa

Other Versions: (COVAR)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: August, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

CPLETH: Choropleth Mapping

Produces maps on the computer line printer using varying shades of darkness to indicate the intensity of phenomena in different geographical areas.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: October, 1971
Similar Programs: CMAP, XMAPP, COBMAP, INTRMAP
Contributor: Department of Geography
University of Michigan

CURVES: Plotting Series of Superimposed Curves

Program for plotting a series of up to four superimposed curves. The pen-and-ink plot is intended to permit visual analysis of the covariation or progression of several different statistical series.

Computer: CDC 6600
Compiler: 6000 FORTRAN EXTENDED
Date Received: March, 1972
Contributor: Department of Geography
University of Minnesota

DENDRO: Dendrograph Plotter Program

The program produces a two dimensional plotter diagram for depiction of the mutual relationships among a group of objects whose pair-wise similarities are given. Displays both within-group and between-group similarities measured as correlation coefficients or distance functions.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

DISAGG: Geographical Matrix Disaggregation

To prepare data from pairs of geographical matrices for input to a multiple regression program.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Contributor: Department of Geography
University of Michigan

Other Versions: (DISAGG)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: December, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

DISGRP: Distance Grouping

Performs a cluster analysis based on Ward's technique of minimum group distance (within group sum of squares) increase at each step.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: November, 1972
Similar Programs: CNGRP, CONGRP
Contributor: Department of Geography
University of Waterloo

DISITR: Discriminant Iterations

Performs an iterative multiple discriminant analysis using orthogonal data.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: May, 1970
Similar Programs: CENDA, COSINE, ITERIM, MDISC, NEARNBR, NOLIDA
Contributor: Department of Geography
University of Iowa

DISTORT: Distortion of Geographical Maps

Calculates the distortion on geographical maps via the theorem of Tissot. The program is for distortion on projections given as a five degree table and uses a finite difference approximation to the derivatives (The program can read punched output from CART2).

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: October, 1971
Contributor: Department of Geography
University of Michigan

DON: Optimal Clustering Routine

Partitions objects into optimally homogeneous groups on the basis of empirical measures of similarity among those objects. Two methods are used to form clusters.

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Similar Programs: CLUSTR
Contributor: Department of Geography
University of California, Los Angeles

DOUBLE: Double Fourier Series

Computes a double fourier series for irregularly spaced data. The series has been used as an alternative model to the polynomial for trend surface analysis.

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Similar Programs: SINGLE
Contributor: Department of Geography
University of California, Los Angeles

DSTAZ: Computation of Distances on a Sphere

Rapid computation of the great circle distance and azimuth between any two points on either the lunar or terrestrial sphere.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: GEODIS, SDIS, SPHERE
Contributor: Department of Geography
Northwestern University

Other Versions: (DSTAZ)

Computer IBM 360/91
Compiler FORTRAN G
Date Received December, 1971
Contributor Department of Geography
University of California, Los Angeles

DYAD: Trend Surfaces by Eigenvector Dyads

The program produces a series of trend surfaces each consecutive member of which better approximates an original input distribution.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Similar Programs: MPLSQ, POLYFIT, TREND, TRENDC
Contributor: Department of Geography
University of Michigan

ELIPS: Plotting of Bivariate Standard Deviations

The program may be used to provide some summary measures of geographical distributions. Includes pen-and-ink plot output.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Similar Programs: CENTRO
Contributor: Department of Geography
University of Michigan

ENTROPY: ntropy Maximizing Techniques

A set of five programs written as subroutines in FORTRAN orresponding to a family of gravity models using entropy maximizing techniques. These spatial interaction models correspond in order to: (1) unconstrained, (2) origin constrained, and (3) total costs constrained flows.

Computer: not stated
Compiler: FORTRAN
Date Received: July, 1972
Contributor: Centre for Land Use and Built Form Studies
University of Cambridge

EQUAL: Lines of Equilibrium

The program uses the pen-and-ink plotter to depict lines of equilibrium about multi-centers of attraction. Such centers could include cities, individual firms, public institutions, etc.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: May, 1970
Similar Programs: GENEB, NAYBOR
Contributor: Department of Geography
University of Iowa

Other Versions: (EQUAL)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: December, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

EUCLID: Geographical Distributions Correspondence in Euclidean Space

Computes the spatial correlation and linear regression between two associated distributions of N points each. The dependent set is in (U,V) space and the independent set is in (X,Y) space.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: October, 1971
Contributor: Department of Geography
University of Michigan

Other Versions: (EUCLID)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: January, 1972
Contributor: Computer Institute for Social Science Research
Michigan State University

EXTRAP: Univariate Geographical Forecasting

Extrapolation of geographical matrices in the time domain using a positionally invariant, time varying, linear, local operator.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Contributor: Department of Geography
University of Michigan

Other Versions: (EXTRAP)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: December, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

FLTREC: Flow-Linkage Trend Recognition

Points or areas are linked to similar neighbors within a search region and these linkages are used to indicate similarities or dependencies. Resulting planar graphs may be useful in indicating the spatial pattern of a process.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

FORFIT: A double Fourier Surface-Fitting Program for Irregularly Spaced Data Points

The double fourier series of trigonometric functions can be used as an alternative model to power-series polynomial trend-surface analysis. It provides a means for interpolation and limited extrapolation of data that are suspected to be periodic.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: DOUBLE, SINGLE
Contributor: Department of Geography
State University of New York, Albany

GENEB: Geographical Neighbors

Given NC points identified by rectangular coordinates X(I), Y(I); I = 1,...NC, the program produces the adjacency matrix of neighbors of order NN less than or equal to 9. These can then be used to find the neighbors to points of a regular lattice.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Similar Programs: EQUAL, NAYBOR
Contributor: Department of Geography
University of Michigan

Other Versions: (GENEB)

Computer.	CDC 6500
Compiler:	6000 FORTRAN EXTENDED
Date Received:	August, 1971
Contributor:	Computer Institute for Social Science Research Michigan State University

GEODIS: Distances from Coordinates

The program reads a vector of geographical coordinates, and then punches the scalar product matrix of all pairs of great circle distances, in degrees.

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	March, 1972
Similar Programs:	DSTAZ, SDIS, SPHERE
Contributor:	Department of Geography University of Michigan

GEOFIT: Determination of Geographical Origins

The program estimates sets of source coordinates from empirical geographical distributions.

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	August, 1970
Contributor:	Department of Geography University of Michigan

GEOPAK: A package of FORTRAN IV programs for the spatial filtering of digital geological maps.

Computer:	IBM 360
Compiler:	FORTRAN G
Date Received:	June, 1974
Similar Programs:	SMOOTH, MAVE
Contributor:	Geosystems division of Lea Associates Limited

GIPSY: Geographical Incremental Plotting System

A computer mapping program for producing point and line symbols on the digital increment plotter.

Computer: IBM 360/50
Compiler: FORTRAN G
Date Received: June, 1971
Similar Programs: MAPIT, POPMAP, SYMBOLS
Contributor: Department of Geography
University of Rhode Island

Other Versions: (GIPSY)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: July, 1972
Contributor: Computer Institute for Social Science Research
Michigan State University

GRAVITY: Iterative Fitting of Gravity Model

Used to fit the social gravity model to data describing spatial interaction (either inflow or outflow) between a set of areas and a user selected focal area.

Computer: CDC 6500
Compiler. 6000 FORTRAN EXTENDED
Date Received: July, 1972
Contributor: Department of Geography
Michigan State University

GRAVTY: Gravity Model Fitting Program

Program involves a data matrix that measures the inflows or outflows from one point to other points (observations). For each observation, distance, mass and index of interaction are calculated. Additional information is provided by subjecting a data matrix to simple regression using index of interaction as dependent variable and mass and distance as independent variables.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: GRAVITY
Contributor: Department of Geography
State University of New York, Albany

GRID: Geographical Interpolation

Interpolation to a square lattice from measures given at scattered geographical (X,Y) positions.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Similar Programs: INTRPOL
Contributor: Department of Geography
University of Michigan

Other Versions: (GRID)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: August, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

GVAR: Scale-Variance Detector with Even Hierarchy

The program uses a modified fixed effects analysis of variance procedure to detect variation at different nested scale levels from rectangularly gridded data.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: March, 1972
Similar Programs: HVAR
Contributor: Department of Geography
University of Michigan

HAAAG: Diffusion on a Regular Lattice

This program is based on the spatial diffusion model of Hagerstrand and simulates the location and number of knowers after each generation for each simulation.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: NONCEL, NONCELO
Contributor: Department of Geography
Northwestern University

HAGMEVA: Spatial Series Summary of HAAAG Output

To provide summary information on the spatial series generated by the tape output version of HAAAG.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: SANDM
Contributor: Department of Geography
Northwestern University

HAGPLOT: Temporal Series Summary of HAAAG Output

To provide summary information on the temporal series generated by the tape output version of HAAAG. Includes pen-and-ink plot output.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: NONPLOT
Contributor: Department of Geography
Northwestern University

HENTRC. Information Theory Entropy Measures

The program computes measures of information theory entropy for bivariate distributions. The following measures are calculated: information content, row entropy, column entropy, expected joint information, information gain, maximum entropies, and certain measures of redundancy.

Computer: IBM 370/135
Compiler: DOS Full FORTRAN
Date Received: October, 1974
Contributor: Department of Geography
University of Ibadan, Nigeria

HEXAGON: Hexagon Plotting and Transformation

Program to draw (using the pen-and-ink plotter) and transform hexagons.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: March, 1972
Contributor: Department of Geography
University of Michigan

HVAR: Scale-Variance Detector with Uneven Hierarchy

The program uses a modified fixed effects analysis of variance to detect scale variation in an uneven fully nested hierarchy.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: March, 1972
Similar Programs: GVAR
Contributor: Department of Geography
University of Michigan

INDIFF: Transaction Flows Analysis

A generalized program for the analysis of transaction flows.

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Similar Programs: ADJUST
Contributor: Department of Geography
University of California, Los Angeles

INPOUT: Simplified Input-Output Model for Interindustry and Interregional Transaction Flows

For each run, the program outputs final demand and production vectors for a two or three region interregional flow model based on user specified input-output coefficients and final demands for each sector.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

INICYL: Interrupted Cylindrical Map Projection

Program to print out the earth's graticule according to the cylindrical projection specified.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: LATMAP, PLOTMAP
Contributor: Department of Geography
State University of New York, Albany

INTPMED: Evaluation of Student Solutions to Location Problems on a Network

The program uses the spatial assignment section of program ALLOC to partition the demand points among the centers chosen by the students. The optimum solution is instructor input and the student's solution is compared to it, giving a ranking of the coefficient of efficiency and a relative frequency distribution.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Contributor: Department of Geography
University of Iowa

INTRMAP: Interactive Choropleth Mapping

This program is an interactive version of CMAP. It is designed to be used on standard character cathode ray tubes with 78 character lines or less.

Computer: IBM 360
Compiler: FORTRAN G
Date Received: November, 1975
Similar Programs: CMAP, COBMAP, CPLETH, XMAPP
Contributor: Department of Geography
Virginia Commonwealth University

INTRPOL: Grid Interpolation

Irregularly spaced data are rendered regular by local interpolation onto a cartesian grid.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: May, 1970
Similar Programs: GRID
Contributor: Department of Geography
University of Iowa

INVERSE: Inverse Bivariate Interpolation

A program to perform inverse bivariate interpolation on a given set of tables of latitude and longitude coordinates for a map projection. A drawing of the resultant inverse tables is made via a pen-and-ink plotter.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: October, 1971
Contributor: Department of Geography
University of Michigan

IOWAP:

The program graphs the cumulative percentile distances on a pen-and-ink plotter using as input the matrix described in output on CONDIST.

Computer: CDC 3600
Compiler: 3600 FORTRAN
Date Received: June, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

IPCIDA: Summarized Areal Classifications

This program is designed to summarize areal classifications based on point samples. It is an aid to characterizing polygon-shaped areas, defined in a larger realm, on the basis of the number of point-occurences of some discrete phenomenon found to be contained in each of these areas. The total weighted values of the data points found in each area are summed and output.

Computer: IBM 370
Compiler: FORTRAN G
Date Received: May 1976
Contributor: Department of Geography
University of Denver

ISODEN: Digital Isodensitometry

The program uses the isodensitracer technique for the automatic mapping of isopleths.

Computer: CDC 6000 series (model not specified)
Compiler: FORTRAN
Date Received: January, 1976
Contributor: Department of Geography
Tel-Aviv University

ISOMET: Isometric Diagram Plotting

The program acts as a driver program that can access any one or all of the (13D) three dimensional isometric drafting subroutines. These subroutines provide an efficient and fast method of displaying a surface--- a single valued function or two variables--in isometric or similar pictorial representation.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Similar Programs: BLCK, MAP3D, PERS, SURGE, TRID
Contributor: Department of Geography
Northwestern University

ITERIM: Iterative Improvements Program for Evaluation and Improvement of Classifications

The program is designed primarily to assess and improve classifications, although it can be used also for principal components analysis, discriminant analysis, and one-way multivariate analysis of variance. Three criteria, pooled within groups sums of squares, Wilks' Lambas, and the sum of the eigenvalues associated with discriminant functions, are computed to assess and compare classifications.

Computer:	CDC 6400
Compiler:	6000 RUN FORTRAN
Date Received:	December, 1971
Similar Programs:	CENDA, COSINE, DISITR, MDISC, NEARNBR, NOLIDA
Contributor:	Department of Geography Northwestern University

KCOLOR: Contiguity Measures for K-Color Maps

Computation of statistics for contiguity measures to provide an evaluation of the hypothesis of randomness in the areal arrangement of the values of a specified property.

Computer:	CDC 6400
Compiler:	6000 RUN FORTRAN
Date Received:	June, 1970
Similar Programs:	CONRAT, VALRAT, VALRATI
Contributor:	Department of Geography Northwestern University

KOPPEN: Koppen Climatic Classification

Classified climatic stations into major climatic types according to standard Koppen criteria.

Computer:	IBM 360/91
Compiler:	FORTRAN G
Date Received:	December, 1971
Similar Programs:	CLIMAT
Contributor:	Department of Geography University of California, Los Angeles

LANDUSE: Land Use and Market Area Model

Development of land use and market area patterns according to the basic Thunen model.

Computer:	CDC 6400
Compiler:	6000 RUN FORTRAN
Date Received:	December, 1971
Similar Programs:	SOLUP, SIRES
Contributor:	Department of Geography Northwestern University

Other Versions: (LANDUSE)

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: October, 1972
Contributor: Department Of Geography
University of Michigan

LAP: Location Allocation Package

The program solves constrained and unconstrained continuous space location-allocation problems by variations of the alternation heuristic (algorithm). Transportation and Weber problems can be solved by requesting zero iterations

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Similar Programs: ALTERN, MAPTRANS, TORN, WEBER, WEBERI
Contributor: Department of Geography
University of Iowa

LATMAP: Mapping Latitude/Longitude Coordinates on Polyconic Projection

Converts latitude/longitude points to Cartesian coordinates based on a polyconic projection, and as options: (1) interpolates the coordinates to a lattice, and (2) constructs an isarithmic line printer map of the data.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: December, 1971
Similar Programs: GRID, INTRPOL, RGRID
Contributor: Computer Institute for Social Science Research
Michigan State University

LINEAR: Linear Pattern Recognition

Local lineations within a scatter of points are detected based on the elongation ratio for a standard deviation ellipse rotated so that its major axis trends along the lineation.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: CLUSTER, REGLAR
Contributor: Department of Geography
State University of New York, Albany

LISTER: Lister/Reproducing Punch

Emulates a lister/reproducing punch combination with greater flexibility and much less user effort. The available options include: (1) reordering of the sequence of variables, (2) integer to decimal and decimal to integer conversion, (3) serial numbering of card output, and (4) multiple copies of listed and punched output.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Contributor: Department of Geography
Northwestern University

LOCATE: Ring and Sector Counting

Rapid counting on a ring and sector basis, about one or more base points, of a series of points whose locations are defined in terms of Cartesian Coordinates.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Contributor: Department of Geography
Northwestern University

Other Versions: (LOCATE)

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: May, 1970
Contributor: Department of Geography
University of Iowa

LOCPOT:

An input-output gravity model simulating industrial attraction

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: March, 1975
Contributor: Department of Geography
University of Georgia

LUAM: Land Use Allocation Model (Version 1)

Development of land use allocations according to the Walsh and Grava Land Use allocation model.

Computer: IBM 7040
Compiler: FORTRAN IV
Date Received: June, 1973
Contributor: Department of Town and Country Planning
University of Sydney

MAPDATA: Descriptive Statistics of Map Data

The input is a rectangular array of data, and the program computes the frequency distribution, the mean, variance, and crude moments. The program then aggregates the array to two by two blocks. Properties of these aggregated data may then be computed.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Contributor: Department of Geography
Northwestern University

MAPIT: Map Drawing on a Pen-and-Ink Plotter

The program performs map-drawing on a pen-and-ink plotter. Input is in the form of a two-dimensional X-, Y-coordinate system.

Computer: CDC 3600
Compiler: 3600 FORTRAN
Date Received: June, 1971
Similar Programs: GIPSY, POPMAP, SYMBOLS
Contributor: Computer Institute for Social Science Research
Michigan State University

MAP3D: Three Dimensional (3-D) Block Diagram

A FORTRAN program for plotting three-dimensional surfaces (topographic as well as statistical) as block diagrams.

Computer: CDC 6600
Compiler: 6000 FORTRAN EXTENDED
Date Received: March, 1972
Similar Programs: BLCK, ISOMET, PERS, SURGE, TRID
Contributor: Department of Geography
University of Minnesota

MAPLOT: A Map Plotting Program

To draw lines (streets, boundaries etc.) and to label them and to plot distinguishing marks at coordinate locations.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: October, 1971
Contributor: Department of Geography
University of Michigan

MAPTRAN: Multiple Facility Location for a Continuous Demand Surface

Map transformation algorithm for locating a set of points in a non-homogeneous population density area such that approximately equal populations are in the dirichlet regions of all points.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Similar Programs: ALTERN, LAP, TORN
Contributor: Department of Geography
University of Iowa

MARKOV1: Regular Markov Chains

Analysis of regular markov chains.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June 1970
Similar Programs: MARKOV2
Contributor: Department of Geography
Northwestern University

Other Versions: (MARKOV1)

Computer:	IBM 360/65
Compiler:	FORTRAN G
Date Received:	May, 1970
Contributor:	Department of Geography University of Iowa

MARKOV2: Simple Absorbing Markov Chains

Analysis of simple absorbing markov chains.

Computer:	CDC 6400
Compiler:	6000 RUN FORTRAN
Date Received:	June, 1970
Similar Programs:	MARKOV1
Contributor:	Department of Geography Northwestern University

Other Versions: (MARKOV2)

Computer:	IBM 360/65
Compiler:	FORTRAN G
Date Received:	May, 1970
Contributor:	Department of Geography University of Iowa

MAVE: Binomially Weighted Smoothing

The program removes high frequency spatial components from a matrix of geographical data by use of a nine point binomially weighted local smoothing operation.

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	August, 1970
Similar Programs:	SMOOTH, SMOTHR
Contributor:	Department of Geography University of Michigan

MDISC: Linear Discriminant Analysis with Classification

A linear discriminant analysis with classification of subjects, based on likelihood ratios, assuming normal distribution and equal variances-covariances.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: November, 1972
Similar Programs: CENDA, DOSINE, DISITR, ITERIM, NEARNBR, NOLIDA
Contributor: Department of Geography
University of Waterloo

MINIMAP: Choropleth Mapping

Designed to produce choropleth maps from map and data file input utilizing small computers.

Computer: IBM 360
Compiler: FORTRAN IV
Date Received: June, 1973
Similar Programs: CMAP, CPLETH
Contributor: Department of Geography and Environmental Planning
Towson State College

MINPATH: Minimum Path Networks

Computation of minimum path networks from each of up to 1000 nodes to all other nodes in a given transportation system. The maximum number of edges is 2000, but multiple jobs may be processed.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: SPA, SPAN
Contributor: Department of Geography
Northwestern University

MLNEGBI: Maximum Likelihood Parameters of Negative Binomial Distribution

This program calculates the maximum likelihood parameters P and K of the negative binomial distribution, from the observed frequency distribution. Using P and K it then fits the negative binomial to the data and computes the chi-square values.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: NEGBIN
Contributor: Department of Geography
Northwestern University

MOCON: Central Moments, Skewness and Kurtosis from Crude Moments

To compute central moments, skewness, and kurtosis from curde moments.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Contributor: Department of Geography
Northwestern University

MOMENS: Computation and Mapping of Areal Moment Surfaces

The program calculates first, second or inverse first areal moments about population masses and optionally produces a line printer contour plot of resulting surface.

Computer: IBM 7040
Compiler: FORTRAN IV
Date Received: June, 1973
Contributor: Department of Town and Country Planning
University of Sydney

MONA: Mapping of Nominally Classified Activitie.

Produces choropleth-type maps of activity patterns from nominally classified activities.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: October, 1971
Contributor: Department of Geography
University of Iowa

MPLSQ: Surface-Fitting Program for Areally-Distributed Data from the Earth Sciences and Remote Sensing

Provides a method of reducing data recorded for a large number of areally-distributed sample sites to the form of simple isopleth maps using various techniques of trend surface analysis.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Similar Programs: DYAD, POLYFIT, TREND, TRENDC.
Contributor: Department of Geography
Northwestern University

MULTI: Multi-Source Location-Allocation Algorithm (Exact)

This program uses a minimum-storage (Backtrack) branch and bound technique to optimally locate M sources to serve N destinations in a Euclidean plane.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Similar Programs: TWAIN, WEBER, WEBERI
Contributor: Department of Geography
University of Iowa

MUNKRE: Transportation Problem

Uses Munkree's Algorithm to solve the transportation problem of linear programming.

Computer: CDC 6500
Compiler: CDC FORTRAN EXTENDED
Date Received: November, 1971
Similar Programs: TRANS
Contributor: Computer Institute for Social Science Research
Michigan State University

NAYBOR: Nearest Neighbor Statistic

Calculation of the nearest neighbor statistic for a set of points distributed in a defined area on a plane. The program may also be used to identify nearest neighbors of each point.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: March, 1972
Similar Programs: EQUAL, GENEB
Contributor: Department of Geography
University of Michigan

NEARNBR: Nearest-Neighbor Discriminant Analysis

Program to classify subjects on the basis of distance to nearest-neighbor individual

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: November, 1972
Similar Programs: CENDA, COSINE, DISITR, INTERIM, MDISC, NOLIDA
Contributor: Department of Geography
University of Waterloo

NEGBIN: Negative Binomial Probability Law

The program computes the individual terms of the negative binomial probability law.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: MLNEGBIN
Contributor: Department of Geography
Northwestern University

NODAC: Node Accessibility Indices

Computation of node accessibility indices (as defined by Shimbel and Katz) for transport networks.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Contributor: Department of Geography
Northwestern University

NOLIDA: Generalized Non-linear Discriminant Analysis with Classification

Computes non-linear discriminant functions which can then be used to classify unknown individuals or to reclassify those individuals from which the discriminant functions were originally derived.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: November, 1972
Similar Programs: CENDA, COSINE, DISITR, ITERIM, MDISC, NEARNBR
Contributor: Department of Geography
University of Waterloo

NONCEL: Simulation of Diffusion through Area of Non-Gridde.. Units

A program for simulation of diffusion through an area of non-grid units, such as political areas and/or other minor civil divisions.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Similar Programs: HAAAG, NONCELO
Contributor: Department of Geography
Northwestern University

Other Versions: (NONCEL)

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Contributor: Department of Geography
University of California, Los Angeles

NONCELO: Simulation of Diffusion through an Area on Non-Grid Units

The overlay of programs permits the simulation of diffusion according to standard Hagerstrand rules over an irregular lattice (non-grid units). The over-lay (1) simulates the location and number of knowers after each generation for each simulation (2) provides summary information on the spatial series generated and (3) provides summary information on the temporal series generated (which includes pen-and-ink plot output).

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: July, 1972
Similar Programs: HAAAG, NONCEL
Contributor: Department of Geography
Michigan State University

NONPLOT: Temporal Series Summary of NONCEL Output

To provide summary information of the temporal series generated on tape output by NONCEL. Includes pen-and-ink plot output.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Similar Programs: HAGPLOT
Contributor: Department of Geography
Northwestern University

NORLOC: Multiple Location Analysis Program

The program solves problem of locating a number of facilities to serve a dispersed population. The routine finds 1) coordinates for positions of facilities, 2) service area of each facility and 3) capacity of each facility from basic input information.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: STDROP
Contributor: Department of Geography
State University of New York, Albany

NORM: Normality Check

The program tests data sets for normality using the Kolmogorov-Smirnov test. If the data are not normal in their original form, the following transformations can be applied: 1) Log, 2) Log-Log, 3) Square Root. The level of significance is chosen by the user.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: May, 1970
Contributor: Department of Geography
University of Iowa

Other Versions: (NORM)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: August, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

NSCA1: Rapid Data Plotting

Rapid data screening. The program produces the N*(N-1)/2 printer plots, linear regressions, and simple correlations for all pairwise combinations of N variables.

Computer: IBM 360/67
Compiler: FORTRAN G
(Requires RUMPLOT subroutines from SHARE library)
Date Received: August, 1970
Similar Programs: REGRESS
Contributor: Department of Geography
University of Michigan

OPTREG: Optimal Regression Analysis

An optimal regression program to calculate regression equations on a dependent variable such that the residual sum of squares is a minimum.

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Contributor: Department of Geography
University of California, Los Angeles

PCPA: Analysis of Paired Comparison Preference Data

A program to perform paired comparisons analysis on preferential data.

Computer: IBM 360/75
Compiler: FORTRAN G
Date Received: March, 1972
Similar Programs: REVPREF
Contributor: Department of Geography
McGill University

PELTO: Pelto D-Function and Relative Entropy

The program calculates two measures of the degree of mixing in multicomponent systems.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Contributor: Department of Geography
University of Michigan

Other Versions: (PELTO)

Computer:	CDC 6500
Compiler:	00 FORTRAN EXTENDED
Date Received:	cember, 1971
Contributor:	omputer Institute for Social Science Research ichigan State University

PERS: Plotting Perspective Views of Surfaces

A set of FORTRAN subroutines that will plot a perspective view of a surface which is represented by a regular grid. Portions of the grid which are hidden from the viewer are not plotted.

Computer:	IBM 360/65
Compiler:	FORTRAN G
Date Received:	January, 1972
Similar Programs:	BLCK, ISOMET, MAP3D, JRGE, TRID
Contributor:	Computing Centre University of British Columbia

PLOTMAP: Map Plotting from Tape

To draw outline maps on various projections to scale from data stored on magnetic tape. Will draw cartograms using double bivariate interpolation.

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	October, 1971
Contributor:	Department of Geogra y University of Michigan

PLOTMP: Printer Grid Mapping Progr

The program plots numerical values and draws an outline around the area of interest

Computer:	UNIVAC 1108
Compiler:	FORTRAN IV
Date Received:	March, 1973
Contributor:	Department of Geography State University of New York, Albany

PLUTS: Special Subroutine for Curve Plotting

A subroutine for time-series plotting on a computer line printer.

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Contributor: Department of Geography
University of California, Los Angeles

POINT: Point-in-Polygon Testing

Used for testing whether a point, defined by Cartesian grid coordinates, is located inside, outside, or on an edge of a polygon defined by a set of coordinates indicating the polygon vertices.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: December, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

POINTS: Point Set Mapped on a Plane or Torus

Computes descriptive measures for a point set mapped onto either a plane or torus. The method of order distance is used to summarize the point pattern.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: November, 1971
Contributor: Department of Geography
Northwestern University

POISSN: Poisson Probability Law

The program computes individual terms of Poisson Probability Law and, as an option, individual terms of the dispersed Poisson Probability Law.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Contributor: Department of Geography
Northwestern University

POL.r IT: Trend Surface Analysis for Degrees One Through Five

Fits algebra olynomials of degree one through five to geog ic data.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Similar Programs: DYAD, MPLSQ, TREND, TRENDC
Contributor: Department of Geography
University of Michigan

POPMAP: Population Maps

The program reads rectangular coordinates (X,Y) and populations, and then draws population maps on the pen-and-ink plotter, with cities represented as circles.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Similar Programs: GIPSY, MAPIT, SYMBOLS
Contributor: Department of Geography
University of Michigan

POPPYR: Population Pyramids

Program to plot population pyramids and compute basic demographic measures.

Computer: 3M 370/158
Compiler: ORTRAN G
Date Received: uly, 1975
Contributor: epartment of Geography
irginia Polytechnic Institute

PREDEN: Dendrograph Input Generation

The program converts raw data into one of a variety of similarity measurements and offers option of punching these as a lower triangular matrix for direct input to the dendrograph program (DENDRO).

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

PRESLOC: Evaluation of Total Costs of a System of Present Warehouses.

The program calculates the costs of the present Warehouse locational configuration for any configuration input by the user. The program may be used in conjunction with program WARELOC.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Contributor: Department of Geography
University of Iowa

PROLO: Probit and Logit Analysis

Permits probit or logit analysis (or both) to be carried out on a data set containing a maximum of 55 variables. The program is basically designed to handle cross-sectional data on which several years of observations may be handled for each member of the cross-section.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Contributor: Department of Geography
Northwestern University

PYRAMID: Population Pyramid Subroutine

A program designed to produce age/sex population pyramids on the computer line printer.

Computer: IBM 360/67
Compiler: FORTRAN WATFIV
Date Received: April, 1972
Contributor: Department of Geography
Pennsylvania State University

RANKD: Ranking Data and Inspecting for Natural Breaks in the Distribution

The program rank orders original data and calculates differences between the ranked numbers and outputs them on a scaled bar graph.

Computer: IVAC 1108
Compiler: DRTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

REGLAR: Regular Pattern Recognition

The program determines the location and areal extent of regions in a point array exhibiting regular, rather than random, distributions. Regularity is assessed in terms of a nearest neighbor test.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: CLUSTER, LINEAR
Contributor: Department of Geography
State University of New York, Albany

REGRESS: Linear Regression Calculation and Plot

A FORTRAN program that calculates and plots (using pen-and-ink plotter) for any two data series, a scatter diagram, the line of regression, and delimits the zone for the standard error of estimate on the scatter diagram.

Computer: CDC 6600
Compiler: 6000 FORTRAN EXTENDED
Date Received: March, 1972
Similar Programs: USCAT
Contributor: Department of Geography
University of Minnesota

REVPREF: Paired Comparisons Analysis from Revealed Spatial Preference Data

Applies the method of paired comparisons to data describing spatial choices of individuals. It develops a proximity matrix which describes the aggregate perceived similarity between locational types.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: June, 1971
Similar Programs: PCPA
Contributor: Computer Institute for Social Science Research
Michigan State University

Other Versions: (REVPREF)

Computer:	IBM 360/65
Compiler:	FORTRAN G
Date Received:	October, 1971
Contributor:	Department of Geography University of Iowa

RGRID: Map Plotting and Contouring

The program produces isamithmic maps on the line printer from scattered observations by first interpolating to a lattice and then contouring these regular values. A location map of the observations, a list of the values at the grid intersections, and the countoured map are returned.

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	August, 1970
Similar Programs:	LATMAP, RGRID2
Contributor:	Department of Geography University of Michigan

Other Versions: (RGRID)

Computer:	CDC 6500
Compiler:	FORTRAN EXTENDED
Date Received:	June, 1971
Contributor:	Computer Institute for Social Science Research Michigan State University

RGRID2: Map Plotting and Contouring

Similar to RGRID; however, this program allows up to eight variables to be mapped.

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	March, 1972
Similar Programs:	LATMAP, GRID
Contributor:	Department of Geography University of Michigan

RIDGREG: Ridge Regression

The program performs ridge regression analysis for a dependent variable and up to 100 independent variables; the data input must be normalized.

Computer: IBM 360/75, 370/165
Compiler: FORTRAN G
Date Received: March, 1973
Contributor: Department of Geography
University of North Carolina

RSTAR: Measure Degree of Association Between Spatially Distinct Observations

This program is designed to measure either the degree to which one retail function (attribute) is spatially associated with any other or the degree to which one shopping center (observation) is like any other.

Computer: IBM 370
Compiler: FORTRAN G
Date Received: May, 1976
Contributor: Department of Geography
University of Denver

SAMPLE: Simulation of Random Sample

Simulation of random sampling with binomial distribution and normal distribution probability tests.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

SANDM: Spatial Series Summary of NONCEL Output

This program reads the output tape produced by the spatial simulation program NONCEL and produces cell means and variances for each generation of each simulation.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Similar Programs: HAGMEVA
Contributor: Department of Geography
Northwestern University

SCALE: Solution of the Law of Categorical Judgement

Transforms original data from the psychological scaling technique known as the method of successive categories to interval scale values.

Computer:
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Contributor: Department of Geography
Northwestern University

SDIS: Spherical Distances

Computes a matrix of distances from a set of points defined by their latitude/longitude coordinates.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Similar Programs: DSTAZ, GEODIS, SPHERE
Contributor: Department of Geography
University of Michigan

SINGLE: Single Fourier Series

Analyzes data that may be represented by curves. The following expression is used for the single fourier infinite;

Y = (A(0)/2) + SIGMA(N) [A(N) COS(N*PI*X)/L + B(N) SIN(N*PI*X)/L]

where N = 1,2,...,infinity.

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Similar Programs: DOUBLE
Contributor: Department of Geography
University of California, Los Angeles

SIRES: Simulation of Urban Residential Segregation

This program uses Monte Carlo simulation procedures to explore the phenomenom of urban residential segregation. The program is designed as a teaching device in that the student can vary a number of parameters such as initial segregation pattern, size and number of neighborhoods to be analyzed, white and black discrimination criteria for moving into or out of a neighborhood, etc. The CDC version is strictly a batch oriented program, but the IBM version is designed for batch or interactive use.

Computer: CDC 6500
Compiler: FORTRAN EXTENDED
Date Received: July, 1975
Similar Programs: LANDUSE, SOLUP
Contributor: Department of Geography
Michigan State University

Other Versions: (SIRES)

Computer: IBM 370/155
Compiler: FORTRAN G
Date Received: January, 1976
Contributor: Department of Geography
Virginia Polytechnic Institute and State University

SLOPE: Spatial Derivative Program

The program calculates the absolute value of the gradient (slope) for data given as a geographical matrix.

Computer: IBM 360/67
Compiler: FORTRAN G
Date Received: August, 1970
Contributor: Department of Geography
University of Michigan

Other Versions: (SLOPE)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: December, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

SMOOTH: Binomially Weighted Smoothing

Removes high frequency spatial components from a matrix of geographical data by use of a nine point binomially weighted local smoothing operation.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: December, 1971
Similar Programs: MAVE, SMOTHR
Contributor: Department of Geography
Northwestern University

SMOTHR: Matrix Smoothing

This assembly language subroutine smooths an array of values (assumed to be spatial data) using the method described by Holloway. It is designed to be called as a subroutine in a FORTRAN program.

Computer: IBM 360/370
Compiler: IBM Assembler
Date Received: February, 1977
Similar Programs: MAVE, SMOOTH
Contributor: Department of Geography
University of California, Santa Barbara

SMRATE: Standard Mortality Rates

The program computes standardized mortality or morbidity ratio and rate for areal-administrative units. The computational method employed adjusts crude death rates for variations in age distribution. SMRATE produces relative mortality ratio as a percentage of the national rate. Standard mortality rate is scaled by user.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

SNORT: Sample Normality Testing

The program tests for normality or lognormality through comparison of the sample frequency distribution and the normal frequency distribution of the same mean and standard deviation as the sample.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: NORM
Contributor: Department of Geography
State University of New York, Albany

SOLUP: Simulation of Land Use Patterns

To construct a map on the printer on which the land use activity which would earn the highest economic rent is computed and identified for each point location. Economic rent is computed from the formula:

R = E(P-A) - E*F*K.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: June, 1971
Similar Programs: LANDUSE, SIRES
Contributor: Computer Institute for Social Science Research
Michigan State University

Other Versions: (SOLUP)

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: October, 1971
Contributor: Department of Geography
University of Iowa

SPA: Shortest Path Algorithm

A program to determine the shortest paths through a network (1) between all pairs of nodes, or (2) between a specified subset of nodes and all other nodes for specified route nodes.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1972
Similar Programs: MINPATH, SPAN
Contributor: Department of Geography
University of Iowa

SPACE:

Computes descriptive statistics for an areal distribution with optional normal probability and/or normal correlation surface tests.

Computer: IBM 7040
Compiler: FORTRAN IV
Date Received: June, 1973
Contributor: Department of Town and Country Planning
University of Sydney

SPAN: Shortest Path Analysis of Networks

Given a network of points interconnected by lines of specified lengths, the program finds the shortest paths between specified pairs of points.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
6000 COMPASS
Date Received: June, 1971
Similar Programs: MINPATH, SPA
Contributor: Computer Institute for Social Science Research
Michigan State University

SPECTR: Computation of Two-dimensional Power Spectra

Computes autocorrelation functions and two-dimensional power spectra for the investigation of the frequency components or wave forms present in an undulating surface.

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Contributor: Department of Geography
University of California, Los Angeles

SPHERE: Spherical Distances

Reads a vector of latitude/longitude coordinates and produces the complete distance matrix for all points.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: August, 1971
Similar Programs: DSTAZ, GEODIS, SDIS
Contributor: Computer Institute for Social Science Research
Michigan State University

STDROP: Algorithm for Dropping Retail Establishments Failing to meet threshold criterion.

Program allocates population to centers, calculates the sales of centers, and discovers whether or not the sales meet the given threshold criterion. If not, the center is either dropped or transformed to a lower central class.

Computer: Honeywell
Compiler: FORTRAN V
Date Received: August, 1972
Contributor: Center for Urban and Regional Studies
Helsinki University of Technology

SURGE: Surfaces from Geographic Grid Data

The program plots three-dimensional surfaces from geographic grid data using the pen-and-ink plotter.

Computer: IBM 360/50
Compiler: FORTRAN G
Date Received: June, 1971
Similar Programs: BLCK, ISOMET, MAP3D, PERS, TRID
Contributor: Department of Geography
University of Rhode Island

Other Versions: (SURGE)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: December, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

SYMBOLS: Plot of Graduated Symbols Map

A FORTRAN program for pen-and-ink plotting of graduated symbols maps where the area of each symbol is proportional to the data value it represents.

Computer:	CDC 6600
Compiler:	6000 FORTRAN EXTENDED
Date Received:	March, 1972
Similar Programs:	GIPSY, MAPIT, POPMAP
Contributor:	Department of Geography University of Minnesota

TAXON: [G]oups Pairs of Similar Operational Taxonomic Units (OTU) [t]o Form Clusters or Nested Hierarchies

Value lies in determining the distances, in N-Dimensional space, between one (OTU) and its next most similar (OTU). Option to group only those (OTU)'s which are spatially contiguous.

Computer:	UNIVAC 1108
Compiler:	FORTRAN IV
Date Received:	March, 1973
Similar Programs:	CNGRP, CONGRP, DISGRP
Contributor:	Department of Geography State University of New York, Albany

THIESEN: Thiessen Polygons

Calculates vertices of Thiessen polygons for a set of points in two-space and then plots the polygons using a pen-and-ink plotter.

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	November, 1972
Contributor:	Department of Geography University of Michigan

TORG: Coordinates from Distances

The program reads a matrix of distance[s], and then produces a vector of plane coordinates.

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	August, 1970
Similar Programs:	TORG2
Contributor:	Department of Geography University of Michigan

Other Versions: (TORG)

Computer:	CDC 6500
Compiler:	6000 FORTRAN EXTENDED
Date Received:	August, 1971
Contributor:	Computer Institute for Social Science Research Michigan State University

TORG2: Coordinates from Distances

The program reads a matrix of distances, and then produces a vector of plane coordinates assuming the distances relate N points in a two dimensional Euclidean space.

Computer:	IBM 360/67
Compiler:	FORTRAN G
Date Received:	March, 1972
Similar Programs:	TORG
Contributor:	Department of Geography University of Michigan

TORN. Tornqvist Multiple Location Algorithm

The program determines the optimum locations for a number of centers C1, C2, C3,...,Cm with respect to the transportation cost of serving N destinations.

Computer:	IBM 360/65
Compiler:	FORTRAN G
Date Received:	January, 1974
Similar Programs:	ALTERN, LAP, MAPTRAN
Contributor:	Department of Geography University of Iowa

TRANS: Solution of the Transportation Problem

The transportation problem is used to determine the minimum total distance (or cost) and the individual flows involved in moving commodities between M number of supply points and N number of demand points given the distance between all pairs of supply and demand points.

Computer: IBM 360/65
Compiler: FORTRAN G
360/MPS
Date Received: May, 1970
Similar Programs: MUNKRE
Contributor: Department of Geography
University of Iowa

Other Versions: (TRANS)

Computer: CDC 6500
Compiler: CDC FORTRAN EXTENDED
CDC APEX
Date Received: December, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

TREND: Trend Surfaces for Degrees One through Six

This program determines any or all trend surfaces from first through sixth degree. Line printer maps may be produced depicting the original data, the contoured surfaces, and the residuals from the surfaces.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: May, 1970
Similar Programs: DYAD, MPLSQ, POLYFIT, TRENDC
Contributor: Department of Geography
University of Iowa

Other Versions: (TREND)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: August, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

TRENDC: Identification of Spatial Patterns of Distance-Decay Type

A program designed to isolate successive distance-decay patterns (derived in a stepwise fashion) around central points and the residuals from these patterns.

Computer: IBM 360/91
Compiler: FORTRAN G
Date Received: December, 1971
Similar Programs: DYAD, MPLSQ, POLYFIT, TREND
Contributor: Department of Geography
University of California, Los Angeles

TRID: Three-Dimensional Isometric Diagrams

A program for pen-and-ink plotting of three-dimensional surfaces as block diagrams. Portions of the grid hidden from the viewer are not plotted.

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: July, 1972
Similar Programs: BLCK, ISOMET, MAP3D, PERS, SURGE
Contributor: Computer Institute for Social Science Research
Michigan State University

TWAIN: Two-Source Location-Allocation Algorithm

This program generates all geometrically possible two-group partitions of N points in a plane. It optimally solves the two-source location-allocation problem.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Similar Programs: MULTI, WEBER, WEBERI
Contributor: Department of Geography
University of Iowa

USDATA: Procedure for Generating a Data File

36 Socio-economic variables have been tabulated by state and stored internally by data statements. The user can call variables, make computational alternations and add up to 50 newly generated variables. Option is available to have 50 original variables punched or printed out.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Contributor: Department of Geography
State University of New York, Albany

VALRAT: Contiguity Measures

Computation of statistics for contiguity measures to provide an evaluation of the hypothesis of randomness in the areal arrangement of the values of a specified property.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: June, 1970
Similar Programs: CONRAT, KCOLOR, VALRAT1
Contriburor: Department of Geography
Northwestern University

VALRAT1: Contiguity Measures

Similar to VALRAT; however this version includes a test for normality.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: October, 1971
Similar Programs: CONRAT, KCOLOR, VALRAT
Contributor: Department of Geography
University of Iowa

Other Versions: (VALRAT1)

Computer: CDC 6500
Compiler: 6000 FORTRAN EXTENDED
Date Received: December, 1971
Contributor: Computer Institute for Social Science Research
Michigan State University

WARELOC: A Heuristic Program for Locating Warehouses With Facility Costs

A heuristic algorithm to solve for the 'optimal' number, sizes, and locations of warehouses at prespecified nodes in a network, where there are fixed costs associated with the facilities. The program has its own shortest path routine (shimbel method).

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Contributor: Department of Geography
University of Iowa

WEBER: Weberian Weight Table

The program attempts to find the optimum location, for a facility so as to maximize accessibility to or from a number of points, according to a vector equilibrium model that assumes a uniform transport surface.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs LAP, MULTI, TWAIN, WEBERI
Contributor Department of Geography
State University of New York, Albany

WEBERI: One-Source Location Algorithm

This program uses a second order iterative technique to solve for the point of minimum aggregate travel of a set of weighted points in a plane.

Computer: IBM 360/65
Compiler: FORTRAN G
Date Received: January, 1974
Similar Programs: LAP, MULTI, TWAIN, WEBER
Contributor: Department of Geography
University of Iowa

XCUT:

To provide rapid, approximate solutions to the multi-source Weber problem.

Computer: CDC 6400
Compiler: 6000 RUN FORTRAN
Date Received: October, 1974
Contributor: Department of Geography
State University of New York, Buffalo

XMAPP: Choropleth Mapping Using the Line Printer

The program generates character symbol choropleth maps for terminal or line printer output. Data files may be entered by the user, reside in storage, or be system generated when testing map outlines. Several maps of one or more region may be produced in a single run.

Computer: CDC Cyber 72
Compiler: FORTRAN EXTENDED
Date Received: February, 1977
Similar Programs: CMAP, CPLETH, COBMAP, INTRMAP
Contributor: Department of Geography
Eastern Illinois University

ZSCORE: Data Standardization

The program first computes means and standard deviations of variables and then determines Z-scores. Output includes a punched deck of the Z-score matrix.

Computer: UNIVAC 1108
Compiler: FORTRAN IV
Date Received: March, 1973
Similar Programs: NORM
Contributor: Department of Geography
State University of New York, Albany

PROGRAM ORDERS
(revised June, 1977)

The Geography Program Exchange makes available all of its computer program holdings according to any of the following options for program orders:

(1) OPTION 1: Listing of Program Documentation and Listing of Program Source Code (Prices for option 1 orders reflect charges for paper stock, computer time, postage, and handling).

Items included in an option 1 order:

(a) Listing of Program Documentation
--Statement of purpose
--Input requirements and deck make-up
--Description of output
--References
--Program source

(b) Listing of Source Code
--Main program
--Internal functions and subroutines
--Test data set (if available)

(2) OPTION 2: Listing of Program Documentation, Listing of Program Source Code on Punched Cards (Prices for option 2 reflect charges for paper and card stock, computer time, postage, and handling).

Items included in a option 2 order:

(a) Listing of Program Documentation (includes the same features as those listed under item 1-a).

(b) Listing of Source Code (includes the same features as those listed under item 1-b).

(c) Program Source Code on Punched Cards.*
--Main program
--Internal functions and subroutines
--Test data set (if available)

(d) Sample Program Output--if sample output is available for a given program, an asterisk (*) will follow the option 2 price for that program on the price list.

(3) OPTION 3: Program Documentation and Program Source Code on Magnetic Tape (Prices for option 3 orders reflect the charges for computer time and handling.** Charges for paper stock and computer time are added if sample output is available).

(a) Program Documentation (including the same features as those listed under item 1-a)recorded in card-image form on magnetic tape.

(b) Program Source Code (including the same features as those listed under item 1-b) recorded in card-image form on magnetic tape.*

(c) Sample Program Output--If sample output is available for a given program, an asterisk (*) will follow the option 3 price for that program on the price list.

*All program source code is made available by the GPE in either BCD or EBCDIC coding. BCD coding contains the basic character set of the IBM-026 keypunch, while EBCDIC contains the expanded character set of the IBM-029 keypunch. All orders should specify which type of coding is desired. If none is specified, the order will be filled using BCD coding.

In addition, the reader will notice in the following "Index of Programs Currently Available" that the GPE usually maintains programs in either CDC or IBM FORTRAN. Several of the programs, however, are made available in both versions of FORTRAN. It is important, therefore, that all orders indicate which version of FORTRAN is desired for each program requested. In the case where the type of FORTRAN is not specified for a program available in either version, the CDC FORTRAN version will default.

**A base tape charge of $20.00 is incurred with each order of one or more programs recorded on magnetic tape (order Option 3). This price includes the cost of a new magnetic tape (2400 feet at $15.50), insurance and postage ($3.00), and handling ($1.50). All magnetic tapes are insured for $50.00 and mailed by parcel post.

The facilities of the Computer Laboratory at Michigan State University permit the GPE to create either seven-track or nine-track magnetic tapes at densities of 200, 556 (7 track only), 800, or 1600 (9 track only) bits per inch.

PAYMENT POLICY

When ordering program materials from the GPE, please send a check, money order, or purchase order for the total amount of the order. Checks and money orders should be made payable to MICHIGAN STATE UNIVERSITY and forwarded to:

Geography Program Exchange
Computer Institute for Social Science Research
510 Computer Center
Michigan State University
East Lansing, MI 48824

IMPORTANT: Payment must accompany an order totalling less than $20.00. The policy of prepayment is also in effect for all orders, regardless of the total dollar value of the order, originating outside the U.S.A. In addition, all foreign orders must be paid in either U.S. currency (example: bank draft on U.S. bank) or by international money order. All prices noted on the price list of programs currently available from the GPE are subject to change without further notice.

SHIPMENT POLICY

DOMESTIC ORDERS (U.S.A. and Canada): All printed matter will be forwarded to the customer's address through third class mail. All parcels (i.e., magnetic tapes and packages of punched cards) will be shipped via parcel post. If shipment by air express is desired, the GPE will provide a quotation for the air express charge upon special request.

INTERNATIONAL ORDERS: All parcels will be forwarded to the shipping address via international parcel post. If a less time-consuming mode of shipment is desired, the GPE will provide a quotation for that shipment charge (e.g., via air express) upon special request. In the case of international orders, any special shipment charge must be prepaid.

Program Prices and Option Availability

Acronym	IBM Option 1	IBM Option 2	IBM Option 3	CDC Option 1	CDC Option 2	CDC Option 3
ACCESS	N.A.	N.A.	N.A.	$12.40	$22.70	$11.10
ADJUST	N.A.	N.A.	N.A.	9.40	16.80*	10.00*
ALLOC	$ 8.80 #	$13.70 #	$ 7.30 #	N.A.	N.A.	N.A.
ALTERN	8.20 #	12.50 #	6.80 #	N.A.	N.A.	N.A.
AZMAP	8.50	12.30	6.80	8.80	15.30*	8.90*
BASICS	N.A.	N.A.	N.A.	8.70	12.80	6.80
BIVAR(PDP)	8.50	15.70*	9.00*	N.A.	N.A.	N.A.
BLCK	8.20	11.80	6.80	8.50	16.30*	9.30*
CANON	8.70	13.80	6.80	N.A.	N.A.	N.A.
CANTRN(UVAC)	9.50	15.40	8.00	N.A.	N.A.	N.A.
CART1	10.30	17.80	8.50	N.A.	N.A.	N.A.
CART2	9.60	15.20	8.00	N.A.	N.A.	N.A.
CENDA	8.10	11.00	6.80	N.A.	N.A.	N.A.
CENTRO	8.80	13.80	7.30	9.00	19.20*	11.10*
CHIINT	N.A.	N.A.	N.A.	7.80	10.60	6.80
CHOROS	9.80	15.30	8.00	N.A.	N.A.	N.A.
CLCOUNT	8.10	11.00	6.80	8.10	11.10	6.80
CLIMAT(GE)	5.50 #	9.60 #	5.50 #	N.A.	N.A.	N.A.
CLUSTER(UVAC)	8.20	11.40	6.80	N.A.	N.A.	N.A.
CLUSTR	N.A.	N.A.	N.A.	8.50	13.50*	7.60*
CMAP	8.00	10.70	6.80	N.A.	N.A.	N.A.
CNGRP	8.90	13.70	7.30	9.10	24.20*	14.60*
CNTOUR	9.10	13.30	7.30	N.A.	N.A.	N.A.
COBMAP	9.10	13.90	7.30	N.A.	N.A.	N.A.
COLMOG	N.A.	N.A.	N.A.	8.10	11.10	6.80
CONDIST	N.A.	N.A.	N.A.	8.70	12.90	7.30
CONGRP	8.90	14.10	7.30	9.00	13.90	7.30
CONRAT	N.A.	N.A.	N.A.	8.20	11.20	6.80
CONTOUR	N.A.	N.A.	N.A.	8.80	13.30	7.30
CONTR	8.30	12.00	6.80	N.A.	N.A.	N.A.
CONTUR	9.80	16.80	8.50	N.A.	N.A.	N.A.
CONWGT(UVAC)	8.90	13.70	7.30	N.A.	N.A.	N.A.
COORD	8.70	13.30	7.30	N.A.	N.A.	N.A.
CORD(UVAC)	8.70	13.30	7.30	N.A.	N.A.	N.A.
COSINE	8.20	11.10	6.80	N.A.	N.A.	N.A.
COVAR	9.50	16.00	8.00	9.50	19.10*	10.00*
CPLETH	8.70	12.30	6.80	N.A.	N.A.	N.A.
CURVES	N.A.	N.A.	N.A.	8.70	12.50	6.80
DENDRO(UVAC)	9.00	13.70	7.30	N.A.	N.A.	N.A.
DISAGG	8.10	11.00	6.80	8.10	11.00	6.80
DISGRP	9.40	13.50	7.30	N.A.	N.A.	N.A.
DISITR	8.80	13.50	7.30	N.A.	N.A.	N.A.
DISTORT	8.20	11.40	6.80	N.A.	N.A.	N.A.
DON	9.50	15.30	8.00	N.A.	N.A.	N.A.
DOUBLE	8.90	12.60	7.30	N.A.	N.A.	N.A.

Acronym	IBM Option 1	IBM Option 2	IBM Option 3	CDC Option 1	CDC Option 2	CDC Option 3
DSTAZ	$ 8.20	$11.80	$ 6.80	$ 8.20	$11.60	$ 6.80
DYAD	8.70	13.10	6.80	N.A.	N.A.	N.A.
ELIPS	8.20	11.60	6.80	N.A.	N.A.	N.A.
ENTROPY	8.10 #	14.70*#	8.70*#	N.A.	N.A.	N.A.
EQUAL	8.50	12.30	6.80	8.50	14.30	8.20
EUCLID	8.10	11.20	6.80	8.10	12.10*	7.40*
EXTRAP	8.10	11.20	6.80	8.10	10.00	6.80
FLTREC (UVAC)	10.60	13.70	9.20	N.A.	N.A.	N.A.
FORFIT (UVAC)	8.80	12.90	7.30	N.A.	N.A.	N.A.
GENEB	8.50	12.60	6.80	8.50	20.80*	12.60*
GEODIS	8.00	11.10	6.80	N.A.	N.A.	N.A.
GEOFIT	9.50	16.20	8.00	N.A.	N.A.	N.A.
GEOPAK	9.50#	16.00#	8.00#	N.A.	N.A.	N.A.
GIPSY	9.10	13.90	7.30	9.40	17.90	10.70
GRAVITY	N.A.	N.A.	N.A.	8.10	12.60 *	7.60*
GRAVTY (UVAC)	8.30	12.00	6.80	N.A.	N.A.	N.A.
GRID	8.30	12.10	6.80	8.30	13.30*	7.40*
GVAR	8.20	11.80	6.80	N.A.	N.A.	N.A.
HAAAG	N.A.	N.A.	N.A.	8.70	12.50	6.80
HAGMEVA	N.A.	N.A.	N.A.	7.80	10.60	6.80
HAGPLOT	N.A.	N.A.	N.A.	7.80	10.70	6.80
HENTRO	6.80	N.A.	N.A.	N.A.	N.A.	N.A.
HEXAGON	8.50	12.90	6.80	N.A.	N.A.	N.A.
HVAR	8.20	11.60	6.80	N.A.	N.A.	N.A.
INDIFF	8.90	13.50	7.30	N.A.	N.A.	N.A.
INPOUT (UVAC)	8.20	11.20	6.80	N.A.	N.A.	N.A.
INTCYL (UVAC)	9.60	15.80	8.00	N.A.	N.A.	N.A.
INTPMED	8.30 #	13.10 #	6.80 #	N.A.	N.A.	N.A.
INTRMAP	7.80	10.60	6.80	N.A.	N.A.	N.A.
INTRPOL	8.20	11.80	6.80	N.A.	N.A.	N.A.
INVERSE	8.70	13.50	6.80	N.A.	N.A.	N.A.
IOWAP	N.A.	N.A.	N.A.	8.00	10.70	6.80
IPCIDA	8.20	11.10	7.00	N.A.	N.A.	N.A.
ISODEN	N.A.	N.A.	N.A.	6.80	N.A.	N.A.
ISOMET	N.A.	N.A.	N.A.	13.50	24.50	12.30
ITERIM	N.A.	N.A.	N.A.	13.50	21.60	12.30
KCOLOR	N.A.	N.A.	N.A.	8.30	11.80	6.80
KOPPEN	8.80	12.60	7.30	N.A.	N.A.	N.A.
LANDUSE	10.10	16.00	8.50	10.00	16.00	8.50
LAP	9.10 #	16.20 #	7.30 #	N.A.	N.A.	N.A.
LATMAP	N.A.	N.A.	N.A.	9.60	23.10*	13.70*
LINEAR (UVAC)	8.20	11.40	6.80	N.A.	N.A.	N.A.
LISTER	N.A.	N.A.	N.A.	8.30	11.60	6.80
LOCATE	8.50	12.50	6.80	8.50	12.60	6.80
LOCPOT	9.40	18.70*	10.40*	N.A.	N.A.	N.A.
LUAM	3.20	N.A.	N.A.	N.A.	N.A.	N.A.

Acronym	IBM Option 1	IBM Option 2	IBM Option 3	CDC Option 1	CDC Option 2	CDC Option 3
MAPDATA	N.A.	N.A.	N.A.	$ 8.10	$11.00	$ 6.80
MAPIT	N.A.	N.A.	N.A.	10.30	17.50*	9.60*
MAP3D	N.A.	N.A.	N.A.	8.20	11.80	6.80
MAPLOT	$ 8.80	$12.60	$ 7.30	N.A.	N.A.	N.A.
MAPTRAN	8.80 #	14.60 #	7.30 #	N.A.	N.A.	N.A.
MARKOV1	8.70	13.30	6.80	8.50	12.60	6.80
MARKOV2	8.20	12.00	6.80	8.20	11.80	6.80
MAVE	8.00	11.00	6.80	N.A.	N.A.	N.A.
MDISC	9.50	15.30	8.50	N.A.	N.A.	N.A.
MINIMAP	8.80 #	12.10 #	7.30 #	N.A.	N.A.	N.A.
MINPATH	N.A.	N.A.	N.A.	8.30	12.00	6.80
MLNEGBI	N.A.	N.A.	N.A.	8.10	11.10	6.80
MOCON	N.A.	N.A.	N.A.	7.80	10.40	6.80
MOMENS	3.20	N.A.	N.A.	N.A.	N.A.	N.A.
MONA	8.30	11.80	6.80	N.A.	N.A.	N.A.
MPLSQ	N.A.	N.A.	N.A.	13.60	22.20	12.30
MULTI	9.40 #	16.80 #	8.00 #	N.A.	N.A.	N.A.
MUNKRE	N.A.	N.A.	N.A.	8.30	12.80*	7.10*
NAYBOR	8.10	11.20	6.80	N.A.	N.A.	N.A.
NEARNBR	8.00	11.00	6.80	N.A.	N.A.	N.A.
NEGBIN	N.A.	N.A.	N.A.	7.80	10.60	6.80
NODAC	N.A.	N.A.	N.A.	8.70	12.60	6.80
NOLIDA	8.70	12.50	6.80	N.A.	N.A.	N.A.
NONCEL	8.70	12.50	6.80	8.80	12.50	7.30
NONCELO	N.A.	N.A.	N.A.	9.70	25.30*	15.30*
NONPLOT	N.A.	N.A.	N.A.	7.80 #	11.20 #	6.80 #
NORLOC (UVAC)	8.80	12.90	7.30	N.A.	N.A.	N.A.
NORM	8.30	12.00	6.80	8.50	15.30*	8.50*
NSCAT	8.30	12.10	6.80	N.A.	N.A.	N.A.
OPTREG	15.60	21.70	14.10	N.A.	N.A.	N.A.
PCPA	11.30 #	23.60*#	9.80*#	N.A.	N.A.	N.A.
PELTO	8.10	11.40	6.80	7.80	11.40	6.80
PERS	9.00	13.10	7.30	N.A.	N.A.	N.A.
PLOTMAP	10.20	17.00	8.50	N.A.	N.A.	N.A.
PLOTMP (UVAC)	8.20	12.00	6.80	N.A.	N.A.	N.A.
PLOTS	8.00	12.30	7.40	N.A.	N.A.	N.A.
POINT	N.A.	N.A.	N.A.	8.10	12.30*	7.40*
POINTS	N.A.	N.A.	N.A.	8.70	12.00	6.80
POISSN	N.A.	N.A.	N.A.	7.80	10.40	6.80
POLYFIT	10.30	19.10	8.50	N.A.	N.A.	N.A.
POPMAP	8.30	10.60	6.80	N.A.	N.A.	N.A.
POPPYR	7.80	10.60	6.80	N.A.	N.A.	N.A.
PREDEN (UVAC)	8.90	13.80	7.30	N.A.	N.A.	N.A.
PRESLOC	8.10 #	12.00 #	6.80 #	N.A.	N.A.	N.A.
PROLO	N.A.	N.A.	N.A.	11.30	20.60	9.80
PYRAMID	8.10 #	12.00 #	6.80 #	N.A.	N.A.	N.A.
RANKD (UVAC)	8.00	10.70	6.80	N.A.	N.A.	N.A.

Acronym	IBM Option 1	IBM Option 2	IBM Option 3	CDC Option 1	CDC Option 2	CDC Option 3
REGLAR (UVAC)	$ 8.10	$11.00	$ 6.80	N.A.	N.A.	N.A.
REGRESS	N.A.	N.A.	N.A.	$ 8.30	$11.80	$ 6.80
REVPREF	10.30 #	20.10 #	8.50 #	11.70	23.60*	13.40*
RGRID	9.00	14.30	7.30	9.00	16.30*	9.00*
RGRID2	9.10	14.70	7.30	N.A.	N.A.	N.A.
RIDGERE	3.40	N.A.	N.A.	N.A.	N.A.	N.A.
RSTAR	8.80	12.60	7.30	N.A.	N.A.	N.A.
SAMPLE (UVAC)	8.20	12.60	6.80	N.A.	N.A.	N.A.
SANDM	N.A.	N.A.	N.A.	8.00 #	11.60 #	6.80 #
SCALE	N.A.	N.A.	N.A.	10.30	16.00	8.50
SDIS	8.00	11.00	6.80	N.A.	N.A.	N.A.
SINGLE	8.80	12.80	7.30	N.A.	N.A.	N.A.
SIRES	9.40	20.60	11.40	9.00	22.50*	12.70*
SLOPE	8.10	11.10	6.80	8.30	13.30*	7.60*
SMOOTH	N.A.	N.A.	N.A.	8.00	10.70	6.80
SMOTHR	7.80	10.60	6.80	N.A.	N.A.	N.A.
SMRATE (UVAC)	8.10	11.10	6.80	N.A.	N.A.	N.A.
SNORT (UVAC)	9.00	14.10	7.30	N.A.	N.A.	N.A.
SOLUP	8.80	12.90	7.30	8.90	14.10*	8.30*
SPA	9.40 #	16.40 #	8.00 #	N.A.	N.A.	N.A.
SPACE	3.50	N.A.	N.A.	N.A.	N.A.	N.A.
SPAN	N.A.	N.A.	N.A.	8.90	20.60*	12.20*
SPECTR	8.80	12.80	7.30	N.A.	N.A.	N.A.
SPHERE	N.A.	N.A.	N.A.	8.30	13.10	7.30
STDROP (HNYWL)	11.50	21.10	9.80	N.A.	N.A.	N.A.
SURGE	8.80	13.10	7.30	8.70	17.00*	9.80*
SYMBOLS	N.A.	N.A.	N.A.	8.20	11.10	6.80
TAXON (UVAC)	8.20	11.20	6.80	N.A.	N.A.	N.A.
THIESEN	8.70	12.50	6.80	N.A.	N.A.	N.A.
TORG	8.80	14.10	7.30	8.90	14.30	7.30
TORG2	8.80	14.10	7.30	N.A.	N.A.	N.A.
TORN	8.20 #	12.50 #	6.80 #	N.A.	N.A.	N.A.
TRANS	8.10	11.20	6.80	8.10	17.70*	11.40*
TREND	11.30	22.00	9.80	11.50	26.80*	11.10*
TRENDC	10.60	20.20	9.20	N.A.	N.A.	N.A.
TRID	N.A.	N.A.	N.A.	9.00	13.90	7.30
TWAIN	8.20 #	12.60 #	6.80 #	N.A.	N.A.	N.A.
USDATA (UVAC)	9.40	15.30	8.00	N.A.	N.A.	N.A.
VALRAT	N.A.	N.A.	N.A.	8.00	10.70	6.80
VALRAT1	8.50	12.30	6.80	8.70	13.80*	7.60*
WARELOC	8.30 #	13.10 #	6.80 #	N.A.	N.A.	N.A.
WEBER (UVAC)	8.10	11.00	6.80	N.A.	N.A.	N.A.
WEBER1	8.10 #	12.00 #	6.80 #	N.A.	N.A.	N.A.
XCUT	N.A.	N.A.	N.A.	8.80	13.80	7.30
XMAPP	8.00	11.10	7.30	8.00	11.10	7.30
ZSCORE (UVAC)	8.00	11.00	6.80	N.A.	N.A.	N.A.

DIRECTORIO DE ALUMNOS DEL CURSO: "SISTEMAS AVANZADOS DE GEO-INFORMATICA EN PLANEACION", DEL 24 DE JULIO AL 4 DE AGOSTO 1978.

1. SR. JULIO ARGUELLES ROMO ZAMORA Y DUQUE No. 8 COL. TACUBAYA MEXICO 18, D. F.	PEMEX ANALISTA MARINA NACIONAL No. 329 MEXICO 17, D. F. TEL. 515-02-85 EXT. 2488
2. SR. EDUARDO BECERRIL MALDONADO PLUTARCO GONZALEZ No. 411 TOLUCA, MEX. TEL. 450-07	"PARQUE SIERRA MORELOS" ESTUDIOS GEOLOGICOS (CONSERVA_ CION DE SUELOS) CAMARA DE DIPUTADOS DOM. CONOCIDO TOLUCA, MEX. TEL. 574-15 y 593-20
3. SR. ALEJANDRO BUSTOS ARCINIEGA SN. JUAN DE ARAGON No. 269-3 COL. GRANJAS MODERNAS MEXICO 14, D. F. TEL. 781-76-68	ESCUELA DE INGENIERIA DE LA UNIVERSIDAD AUTONOMA DE GRO. CATEDRATICO AV. DE LA JUVENTUD S/N. CHILPANCINGO, GRO. TEL. 2-27-41
4. SR. BULMARO CABRERA RUIZ JURISTAS 6A - C CD. SATELITE TEL. 519-79-39	S.A.H.O.P. JEFE DE LA OFICINA DE FOTOGRAMETRIA AV. XOLA Y UNIVERSIDAD COL. NARVARTE MEXICO 12, D. F. TEL. 519-79-39
5. ARQ. ENRIQUE CAMPOS VEGA MORELOS No. 707 - 13 TOLUCA, MEX. TEL. 536-74	"PARQUE SIERRA MORELOS" JEFE DEL TALLER DE PROYECTOS CAMPAMENTO No. 1 PONTEZUELO TOLUCA, MEX.
6. SR. SERGIO CONSTANTINO PEREZ VALLE SAN JUAN DEL RIO No. 10-2 COL. VALLE DE ARAGON EDO. DE MEXICO	DIRECC. DE PLANEACION TERRIT. DE ASENT. HUMANOS. S.A.H.O.P. ESPECIALISTA URBANO INSURGENTES SUR No. 1443- 2º PISO COL. SAN JOSE INSURGENTES MEXICO 19, D. F. TEL. 563-80-76 - 598-0023-
7. SR. FELIPE DE JESUS CHIU MARTINEZ ANTILLAS No. 308 -18 COL. PORTALES MEXICO 13, D. F. TEL. 672-33-28	COMISION DEL PLAN NACIONAL - HIDRAULICO ANALISTA TEPIC No. 40-3º PISO COL. ROMA SUR MEXICO 7, D. F. TEL. 574-10-73
8. SR. SERGIO DEL VILLAR MARTINEZ AV. UNIVERSIDAD 1953 EDIF.34-304 COL. COPILCO MEXICO 20, D. F. TEL. 550-32-16	FACULTAD DE MEDICINA, UNAM ANALISTA DE SISTEMAS CIUDAD UNIVERSITARIA MEXICO 20, D. F TEL. 548-99-48

9. SR. AGUSTIN FERNANDEZ EGUIARTE
SIEMPREVIVA No. 17
COL. XOTEPINGO.
MEXICO 21, D. F.
TEL. 549-08-99

CENTRO DE CIENCIAS DEL MAR Y LIMNOLOGIA, UNAM.
TECNICO ACADEMICO
CIUDAD UNIVERSITARIA
TEL. 550-52-15 EXT. 4861-4862

10. SR. JAIME FCO. GO MEZ VEGA
XOLA No. 13
COL. DEL VALLE
MEXICO 12, D. F.
TEL. 590-64-82

S.A.H.O.P. SECC. DE ASENT. HUMANOS Y OBRAS PUBLICAS
ANALISTA
AV. UNIVERSIDAD S/N FRENTE A MITLA
COL. NARVARTE
MEXICO 12, D. F.
TEL. 590-64-82

11. SR. ALFONSO GOVELA THOMAE
SOLEDAD No. 28-A
COL. FLORIDA
MEXICO 20, D. F.

UNIVERSIDAD AUTONOMA METRO-POLITANA, XOCH
PROFESOR ASOCIADO
CALZ. DEL HUESO S/N
MEXICO, D. F.

12. SR. JAIME ANDRES HERNANDEZ M.
PITAGORAS No. 763 INT. 3
COL. NARVARTE
MEXICO 12, D. F.
TEL. 523-34-07

AEROPLANIMETRIA VALCA, S. A.
JEFE DEPTO. DE CATASTRO
BLVD. M. A. CAMACHO No. 40 DESP. 208
NAUCALPAN, EDO. DE MEXICO
TEL. 557-65-52

13. SR. FRANCISCO INDALECIO SAÑUDO
NACIONAL No. 3008
COL. 5 DE MAYO
PUEBLA, PUE.

IMPULSORA MEXICANA DE TELCOMUNI-CACIONES, S. A.
SUPERVISOR EN ESTUDIO DEL MERCADO Y PLANEACION
AV. JUARES No. 2318-101
COL. LA PAZ
PUEBLA, PUE.
TEL. 48-54-07 - 48-55-33

14. ING. DAVID LUJAN JOPEZ
EDIF. 82-B-201
COL. COSMOPOLITA
MEXICO 16, D. F.
TEL. 355-16-04

PETROLEOS MEXICANOS
ANALISTA PROGRAMADOR
MARINA NACIONAL No. 329
COL. ANAHUAC
TEL. 250-36-22

15. SR. LUIS EDGARDO LLANES ARENAS
ENRIQUE REBSAMEN No. 27
EDUCADORES, SATELITE
TEL. 562-60-12

CONSULTORES EN PROBLEMAS DEL DEL COMPORTAMIENTO, S. A.
TECNICO EN PROCESAMIENTO DE INFORMACION
HERIBERTO FRIAS No. 1222
COL. DEL VALLE
MEXICO 12, D. F.
TEL. 559-22-22

16. SR. DAVID MATA DORANTES
ORIENTE 166 No. 1160-3
COL. SECTOR POPULAR
MEXICO 13, D. F.
TEL. 582-46-96

BANCO DE MEXICO, S. A. F.I.R.A.
AUXILIAR ANALISTA-AREA DE RIEGO
POR BOMBEO
INS URGENTES SUR No. 2375
MEXICO, D. F.
TEL. 550-70-11

17. SR. ALFONSO NIETO ZERMEÑO
PERIFERICO SUR No. 4421 -104
COL. B. DEL PEDREGAL
MEXICO 21, D. F.
TEL. 568-44-03

FERTIMEX.
INVESTIGADOR
PITAGORAS Y MORENA

18. ANA MARIA OJEDA MUÑOZ
SANCHEZ AZCONA No. 1557-304
COL. DEL VALLE
MEXICO 12, D. F.
TEL. 575-78-94

D.E.T.E.N.A.L.
JEFE DE SECCION DE PROYECTOS
SN. ANTONIO ABAD No. 124
COL. TRANSITO
MEXICO 8, D. F.
TEL. 578-62-00 - 158

19. SR. LUIS JOAQUIN POOT AYALA
DR. VERTIZ No. 757 - 27
COL. NARVARTE.
MEXICO 12, D. F.

DIRECC. GRAL. DE PLANEACION TERRIT.
DE ASENTAMIENTOS HUMANOS
ESPECIALISTA
INSURGENTES SUR No. 1443
COL. SN. JOSE INSURGENTES
MEXICO 19, D. F.
TEL. 598-00-23

20. SR. JOSE RAMOS VALLADOLID
GABRIEL MANCERA No. 322-2
COL. DEL VALLE
MEXICO 12, D. F.
TEL. 536-64-95

SUBDIRECCION DE PROG. AGROP.
Y FORESTAL
JEFE DE INFORMACION Y ESTADISTICA.
FRAY SERVANDO T. DE MIER No. 77-5-PSO
MEXICO 1, D. F.
TEL. 761-24-85

21. SR. OSCAR SOLIS CABALISCO
NUEVA YORK No. 69 - 401
COL. NAPOLES
MEXICO 18, D. F.
TEL. 543-98-64

S.A.H.O.P.
JEFE DE OFICINA
TEL. 598-00-23

22. SR. CARLOS GUILLERMO SUAREZ CASTAÑEDA
5a. PRIVADA DEL CHICLE No. 37
COL. GRANJAS MEXICO
MEXICO 8, D. F.
TEL. 657-13-95

CIA. MINERA AUTLAN, S.A. DE C. V.
ANALISTA
MARIANO ESCOBEDO No. 510 - 4° PISO
COL.POLANCO
MEXICO 5, D. F.
TEL. 250-19-77 EXT. 185

23. MA. ELENA VERDE REDARTE
PLAYA OLA VERDE No. 347
COL. R. IXTLACIHUATL
MEXICO 13, D. F.
TEL. 579-16-67

COMISION DEL PLAN NAL. HIDRAULICO
ANALISTA
TEPIC No. 40
COL. ROMA
MEXICO 7, D. F.
TEL. 584-72-54

24. SR. JOSE GUILLERMO YAÑEZ HEREDIA
AMBATO No. 924-3
COL. LINDAVISTA
MEXICO 14, D. F.
TEL. 586-90-34

COMISION DEL PLAN NAL. HIDRAULI
ANALISIS DE SITEMA
TEPIC No. 40
COL. ROMA
MEXICO 7, D. F.
TEL. 584-72-54

'mg.