Input Methods Panel

Olin Bockes, Presiding
Soil Conservation Service

Dan G. Dixon
DMA Hydrographic Center

Hollis F. Ryan
Calspan Corp.

James O. McDonough
Concord Control Inc.

Peter G. Wohlmut
i/o Metrics

Bockes: As Dr. Boyle said this morning, the structure of this session has been designed to present the current state of the art. "Current", perhaps, depends on where you are in the cycle. If you are a researcher in a university environment where the community is a highly technical, specialized group of people, "current" means one thing. If you operate in a production shop, current technology is a different world, and the people who are working with that technology are a different group of people. I would like to think that everyone, whether in research and development or in production-oriented programs, is as far along as Dr. Boyle suggested. I would like to believe that we are going to have current problems in automated drafting solved in a year or so.

I suspect that our first speaker from "production" will not be quite this optimistic. He is Dan Dixon, a scientific staff assistant at DMA Hydrographic Center. The last panel discussed the LIS system, and Dan is involved in planning and coordinating the implementation, production, and utilization of that system. After Dan, we will hear about some of the newer scanner systems. Panel member Hollis Ryan is from CALSPAN Corp. in Buffalo, N.Y., formerly the Cornell Aeronautical Laboratories. CALSPAN has dealt in image processing for many years and has expanded to line encoding, fingerprint identification, and that sort of work. We also have Jim McDonough from Concord Control and Peter Wohlmut from i/o Metrics, where digitizing by tracking small automic particals was developed.

I think we have come a long way from 5 yr ago when we were talking about etched grooves on metal plates, inertial problems with floating-arm digitizing, and designs for cursors that would facilitate observing, tagging, and locating points.

Dixon: The Defense Mapping Agency Hydrographic Center (DMAHC) may not be exactly a household word to many of you. When DMA was formed several years ago, we were a split off from the Naval Oceanographic Office which at that time had the responsibility for hydrographic charting and mapping, along with various oceanographic functions. Essentially our function is to provide hydrographic charts, publications, and other related products to support both the maritime and naval fleets. This support is provided on a world-wide basis with the exception of those areas encompassing the continental U.S. and its possessions which fall under the cognizance of the National Ocean Survey.

The expansion in the automated cartography program at DMAHC began last year with the acquisition of the Lineal Input System (LIS) which was developed by the Rome Air Development Center under contract to the Planning Research Corp. (PRC). Although LIS represents the first "advanced" equipment for our automated cartographic system, high-precision digital plotting equipment has been in use at DMAHC since the mid 1960's.

Figure 1 illustrates the DMAHC Automated Cartographic System basic functions: data acquisition which LIS currently satisfies; the hydrocartographic data bank, a separate entity; data exploitation which would include the device that Howard Carr mentioned previously--the DIODE or digital input/output display equipment; and graphic finishing which includes various types of high-precision plotting devices, including raster and standard x-y vector plotters. LIS has been on board a year in which a comprehensive 6-mo test and evaluation program was completed including processing actual chart data.

Figure 2 shows the LIS equipment arrangement. This is a typical layout for a work station interfaced to the master processor PDP-15. On the far right is the offline Xynetics proofing plotter. A typical work station includes a Gradicon digitizer interfaced to an Imlac PDS-1G interactive graphics display. The Imlac display shown here is actually mounted on a stand located at one end of the digitizing table. The various work stations function under control of the master processor through a time-sharing arrangement. As various routines are called for via Imlac keyboard interaction, software modules are transmitted from mass storage to the Imlac minicomputer.

This provides for more effective utilization of system processing and data storage/retrieval capabilities, accommodating up to 10 work stations for a single master processor. There are four digitizing work stations now operational; one will be added in Jan. 1975; and five more are tentatively scheduled for acquisition during the next 12 to 16 months.

Figure 3 depicts the present LIS configuration for 4 work stations all served by the PDP-15 master processor. Peripherals include 2 fixed-head disks, 2 disk-pack drives, and other standard peripherals. The Xynetics plotter operates offline and performs quite satisfactorily in generating proof plots of LIS data for editing purposes; it is an extremely high-speed plotter, as fast as 40 in/sec with an overal accuracy of 0.0045 in.

Figure 4 is a shot of the master processor with the disk-pack drives in the middle background and the deck tape units on the far right.

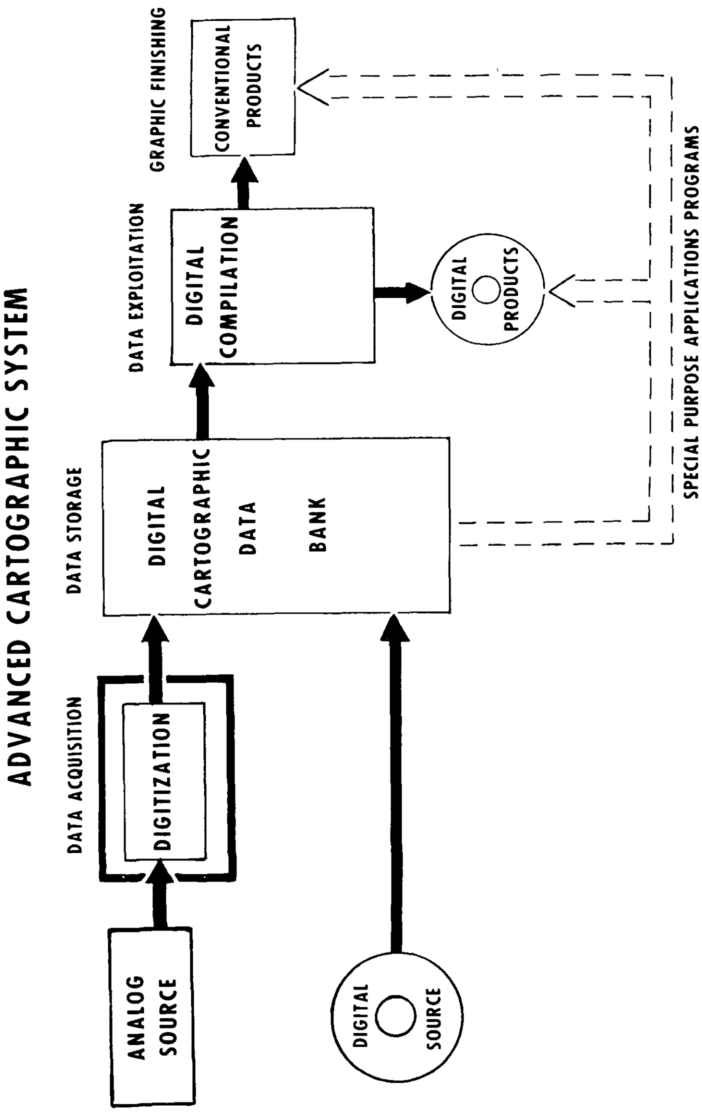

ADVANCED CARTOGRAPHIC SYSTEM
DATA ACQUISITION
DATA STORAGE
DATA EXPLOITATION
GRAPHIC FINISHING
ANALOG SOURCE
DIGITIZATION
DIGITAL CARTOGRAPHIC DATA BANK
DIGITAL COMPILATION
CONVENTIONAL PRODUCTS
DIGITAL PRODUCTS
DIGITAL SOURCE
SPECIAL PURPOSE APPLICATIONS PROGRAMS

Figure 1

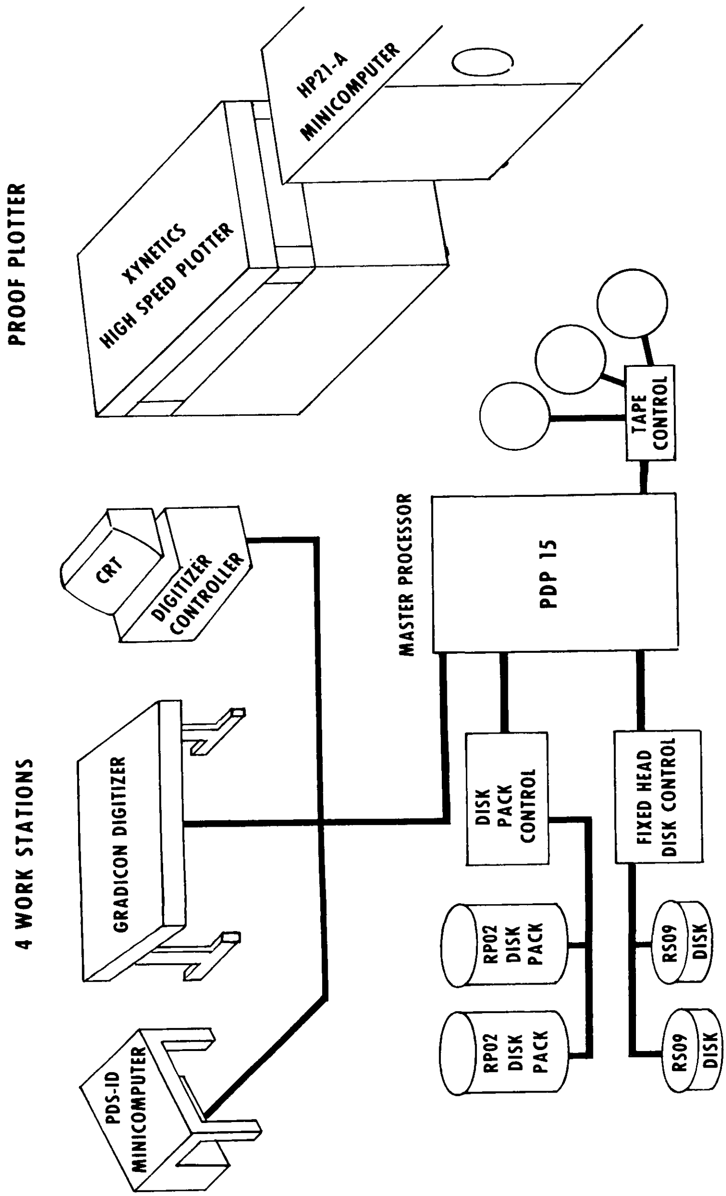

PROOF PLOTTER
XYNETICS
HIGH SPEED PLOTTER
HP21-A
MINICOMPUTER
TAPE
CONTROL
4 WORK STATIONS
CRT
DIGITIZER
CONTROLLER
MASTER PROCESSOR
PDP 15
GRADICON DIGITIZER
DISK
PACK
CONTROL
FIXED HEAD
DISK CONTROL
PDS-ID
MINICOMPUTER
RP02
DISK
PACK
RP02
DISK
PACK
RS09
DISK
RS09
DISK

Figure 2

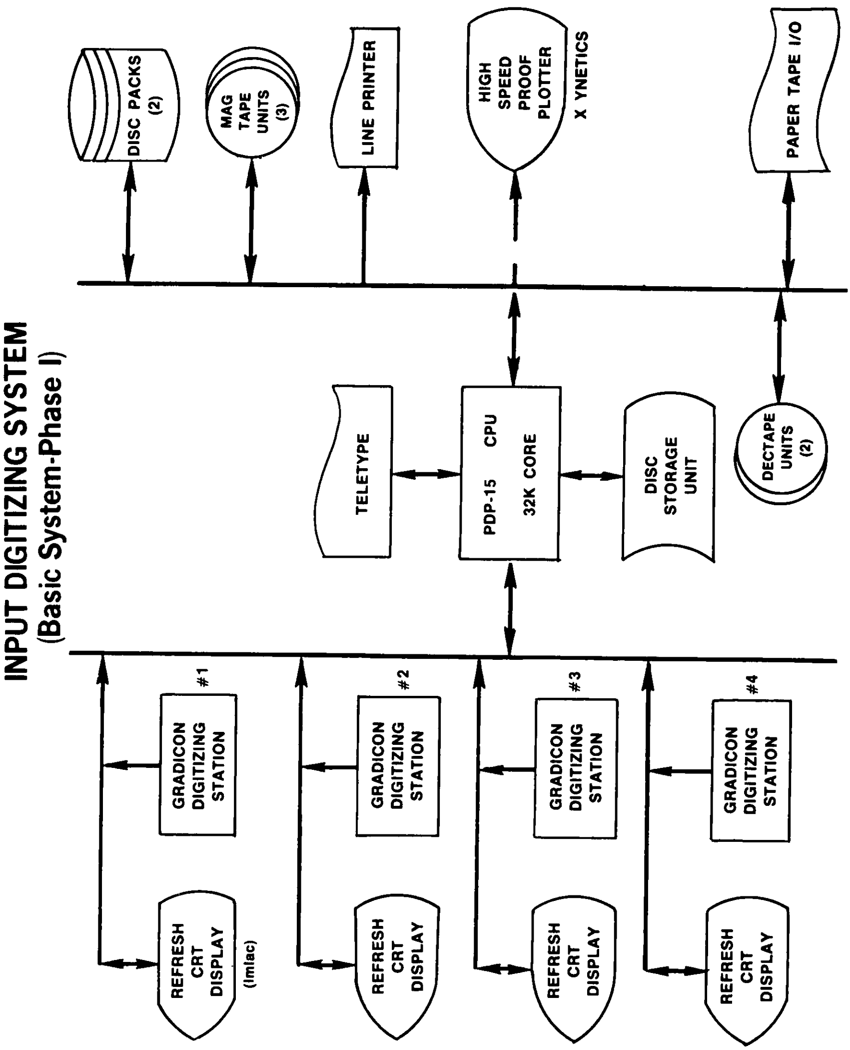

INPUT DIGITIZING SYSTEM
(Basic System-Phase I)
DISC PACKS (2)
MAG TAPE UNITS (3)
LINE PRINTER
HIGH SPEED PROOF PLOTTER
X YNETICS
PAPER TAPE I/O
TELETYPE
PDP-15 CPU
32K CORE
DISC STORAGE UNIT
DECTAPE UNITS (2)
GRADICON DIGITIZING STATION #1
GRADICON DIGITIZING STATION #2
GRADICON DIGITIZING STATION #3
GRADICON DIGITIZING STATION #4
REFRESH CRT DISPLAY (Imlac)
REFRESH CRT DISPLAY
REFRESH CRT DISPLAY
REFRESH CRT DISPLAY

Figure 3

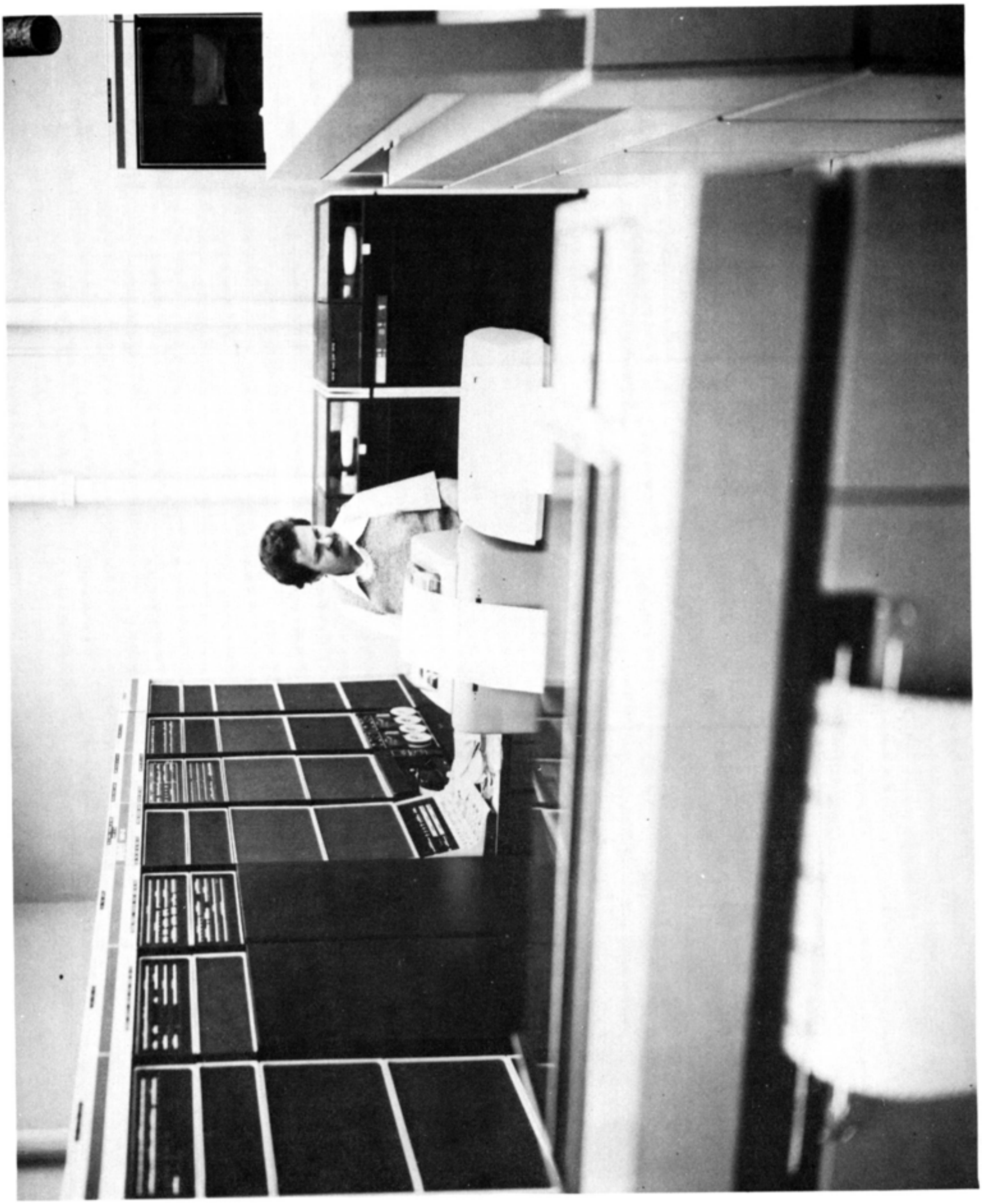


Figure 4

The Xynetics plotter (fig. 5) satisfies the edit/plot-generation function. It is controlled by a Hewlett-Packard 2100 minicomputer with 8-k core which provides significant plot-generation capability for other applications. The plotter is used to generate a number of other interim products such as bathymetric data collection sheets to contain depth or sounding information amounting to 8 k or 9 k values/sheet. Previously, these sheets were generated much slower with a CalComp plotter.

Figure 6 presents a closeup of a work station. Normally, the cartographer performs various functions (create file, file registration, digitize, edit, remote job-entry processing) through sequential responses to queries presented on the display screen. The first procedure (create file) entails entering, via the keyboard, various parameters describing the source to be digitized. When completed, the cartographer would register the new file to the source sheet secured to the digitizing table. At this point, the cartographer could enter the digitization mode and would be presented a list of feature types via the display from which to make a selection.

Figure 7 presents the selection of functions. In order to digitize, the cartographer would punch the number "3". However, before digitizing, the feature must be properly identified or "headered." The cartographer must proceed through a header building scheme in which data would be entered into blank fields for feature class, type, subtype, and 8 associated levels of descriptors. For example, a typical header description for a buoy would include the color, flashing frequency, and ID number. When data bank retrieval and maintenance procedures are fully considered, I think it is essential that a significant description be provided for each feature being digitized for data bank retention and future utilization.

After 2 to 4 hr of digitizing, the cartographer would generate a tape of the digital data and then obtain a Xynetics plot. The cartographer would then register the plot to the source material and identify any discrepancies that might exist. When completed, the cartographer could then select any of the edit procedures available (fig.8) to correct the glitches. These corrections could include deleting a feature, changing the midsection of a feature, or changing the feature header. I sometimes like to refer to this system as "cartographer-proof." Although I don't particularly care for the term, it does relate the significant amount of built-in safeguards. Considering the complexity of the operations of sophisticated systems such as LIS, I don't feel that the safeguards represent needless "over-engineering" at this point. Once we become more familiar with automated operations, I personally feel that many of the safeguards can be eliminated, alleviating many redundant and time-consuming procedures. But for the time being, the present level of system safeguarding is considered necessary.

Once editing has been completed and all changes made, the operator would select a remote job-entry processing routine (fig. 9) for the desired reference frame, such as geographic coordinates. LIS is based on the storage of data by geographic coordinates. The

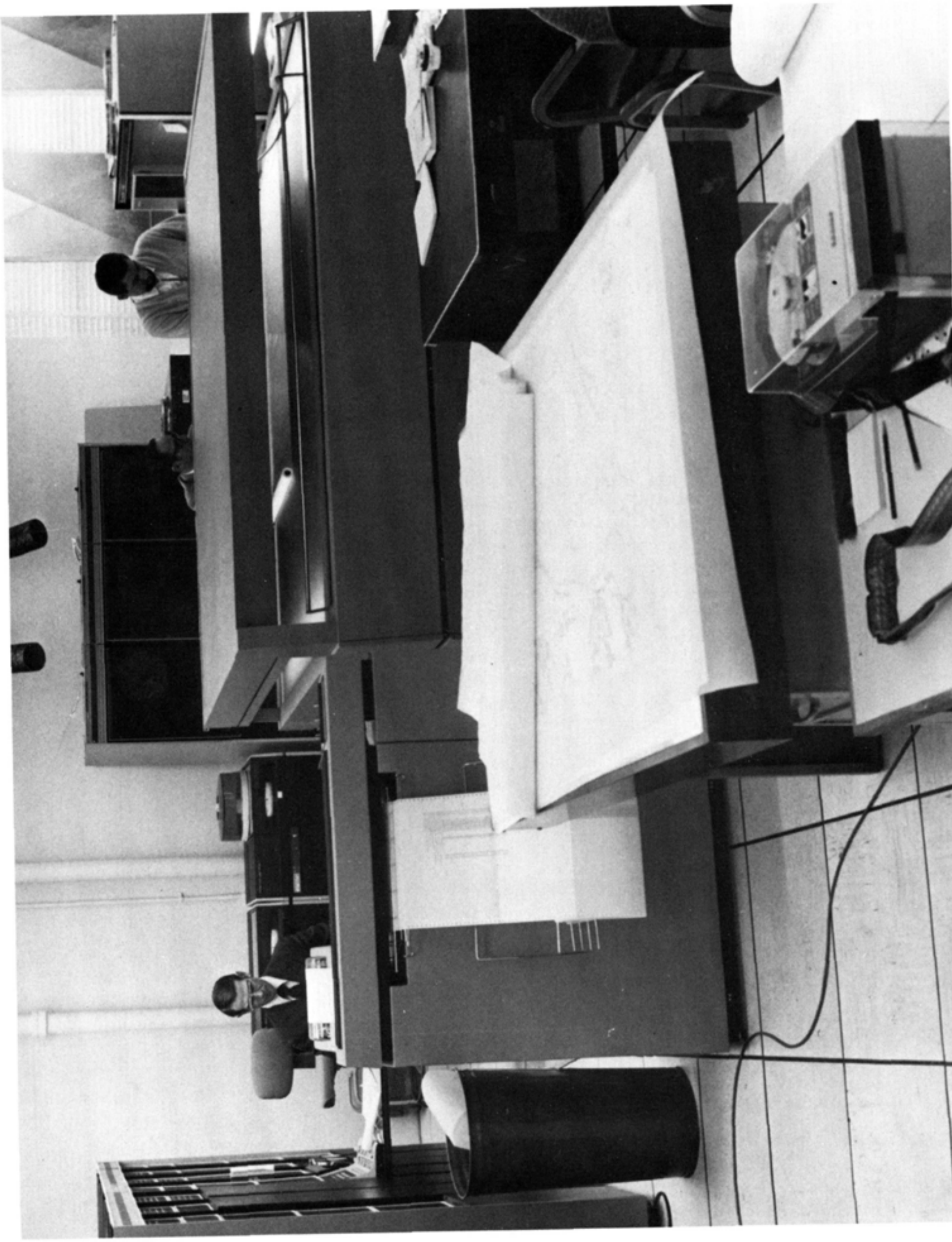


Figure 5

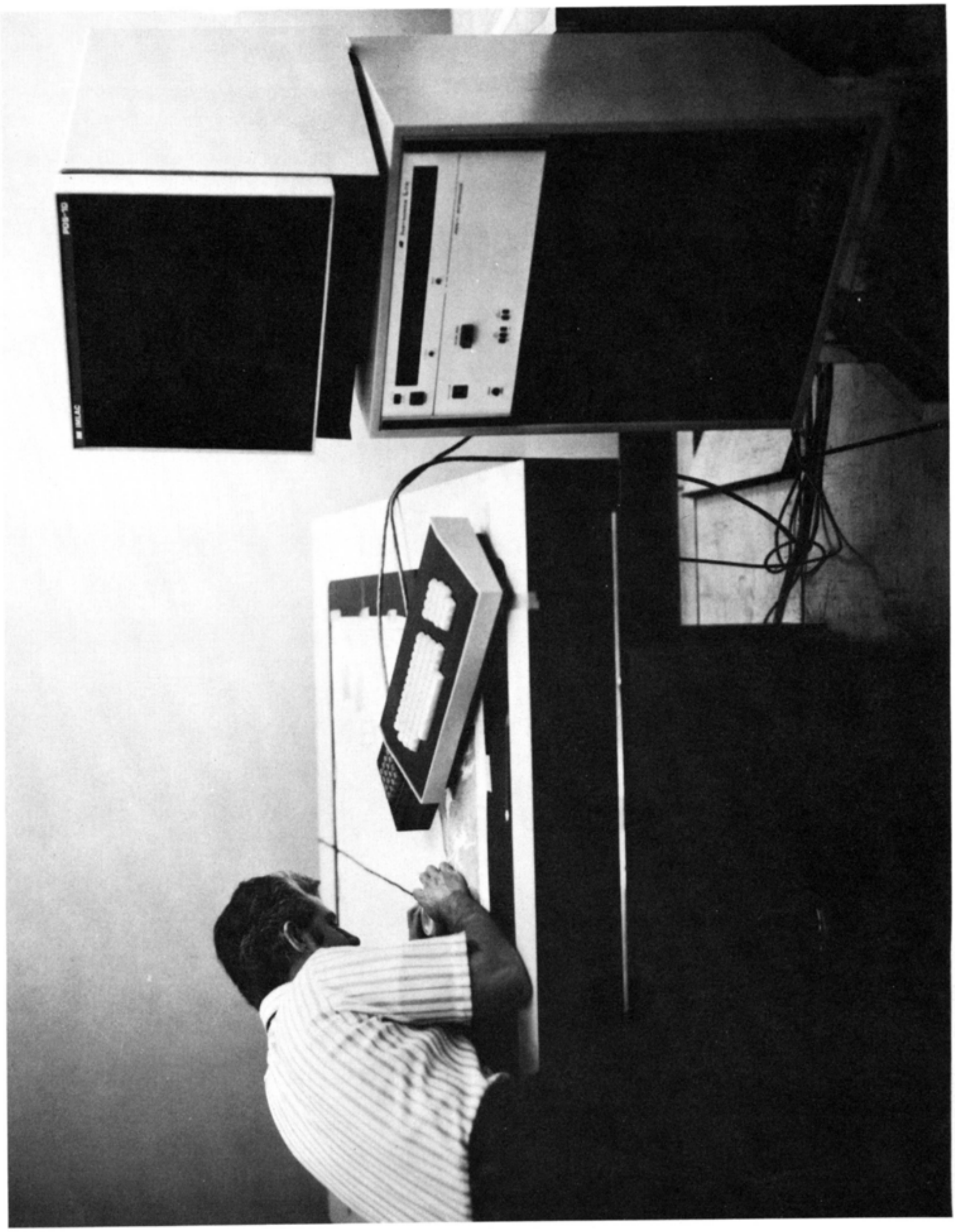


Figure 6

MASTER MODE

1 CREATE FILE

2 REGISTER

3 DIGITIZE

4 EDIT

5 REMOTE JOB ENTRY

6 LOG OFF

ENTER SELECTION NUMBER 3

Figure 7

SPECIAL FUNCTION KEYBOARD

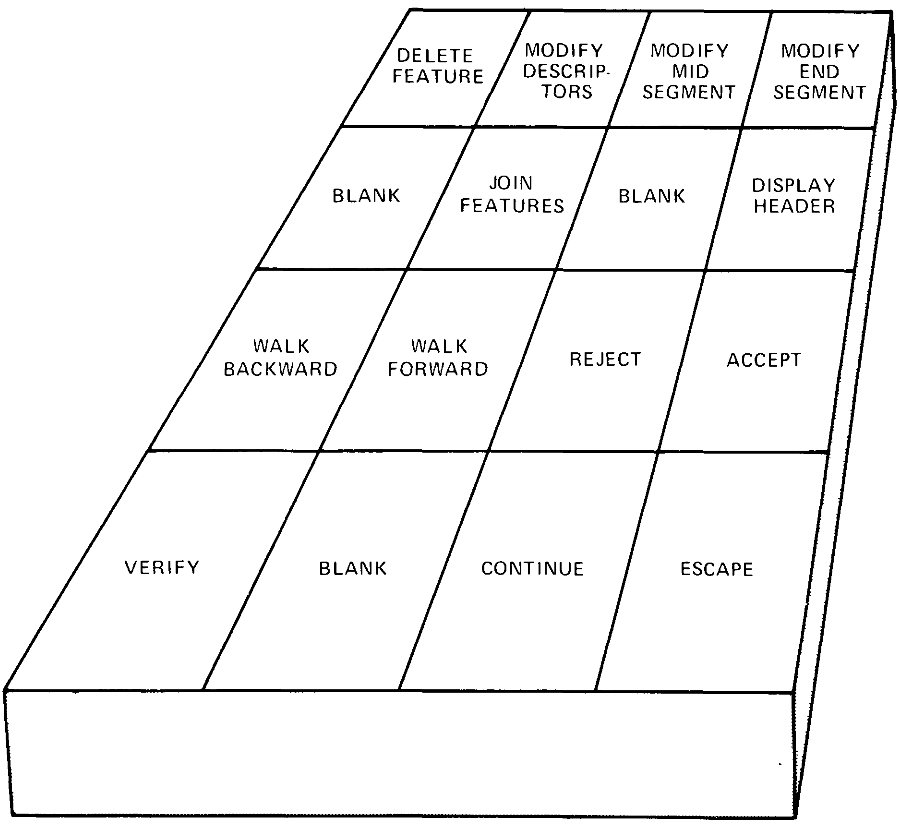


Figure 8

RJE
REMOTE TASK SELECTION

1 INPUT FROM A MAGNETIC TAPE

2 OUTPUT TO A MAGNETIC TAPE

3 TRANSFORM TO GEOGRAPHICS

4 PLOT

5 SECTION GEOGRAPHIC

6 SECTION TABLE

7 CLIP/JOIN

10 PANEL

11 END OF JOB

12 CANCEL JOB

ENTER SELECTION NUMBER

Figure 9

type of data that we are producing with LIS essentially addresses production of two distinctly different items--hard copy and digital products (digital data bases and digital charts). LIS actually provides input to the digital compilation process from which final hard copy is generated. On the other hand, LIS is the primary system from which final digital products will be developed.

LIS is now being used primarily to generate a special-purpose digital data base for the Norfolk, Va., area (fig. 10). Data are being extracted from about 80 charts and maps--70 USGS maps at 1:24,000 scale and 10 NOS charts ranging from 1:20,000 to 1:40,000 scale. Culture, shoreline, and topography are being extracted from the USGS maps, and the NOS charts provide navigational information.

Some rather interesting problems have been encountered. It had been decided that shoreline would be derived from 1:20,000-scale NOS charts and matched with culture and contours from the USGS maps. I am not sure what happened, but when the chart was reduced to 1:24,000 and layed over the contour and culture plots appropriately scaled from the maps, the shoreline did not fit the contours and culture. Roads that were supposed to intersect the middle of piers were in the water along with a number of contours. This is not to say that the NOS charts were less accurate than the USGS maps or vice versa. I think what really happened was that the degree of generalization by the cartographers might have differed enough to result in distinctly different products, and I feel this is not going to be the last time that we will encounter the problem.

As I said, we are presently producing a digital data base with LIS; however, during the next fiscal year data will be digitized to produce special-purpose, small-scale, fleet-oriented graphics. We do not have full capability required, but are aiming for standard nautical chart production. The graphic finishing process alone is a monumental task, particularly automated symbolization for a wide variety of features. Within the next 2 or 3 yr of R&D efforts with Engineering Topographic Laboratories and Rome Air Development Center, we should see significant progress. The ultimate goal is to automate the entire chart production and maintenance program, and we have taken the first steps.

Ryan: I would like to describe how image processing technology has been applied to the digitizing operation. As specialists in image processing, we saw that our techniques could be applied to the map digitizing problem.

Figure 1 lists some applications of image processing technology related to map digitizing. More specifically we have applied the technique of developing algorithms using interactive minicomputer techniques, the image processing algorithms themselves, and the digital hardware implementation technology. There are two programs that I will describe. One refers to the first application, called SEASCAPE, and is being used by the Office of Naval Research for digitizing bathymetry charts. The other program, called DIGIMAP, is being developed on company funds to provide the service of digitizing charts and manipulating the encoded data; the most important manipulation is computer-compositing map sheets.

GEOGRAPHICAL TEST AREA FOR PROTOTYPE NTEC SHIPBOARD RADAR NAVIGATION TRAINER SIMULATOR

LEGEND

AREA-4096 SQ. N.M.
LOCATION-NORFOLK VA. AREA
BASE SCALE-1:24,000
BASIC SOURCE-USGS 7½
MINUTE QUAD SHEETS
NOS HARBOR CHARTS
RADAR SCOPE PHOTOGRAPHY
RECORDING RESOLUTION-40 MICRONS
GROUND RESOLUTION-1½ METERS AT 1:25,000

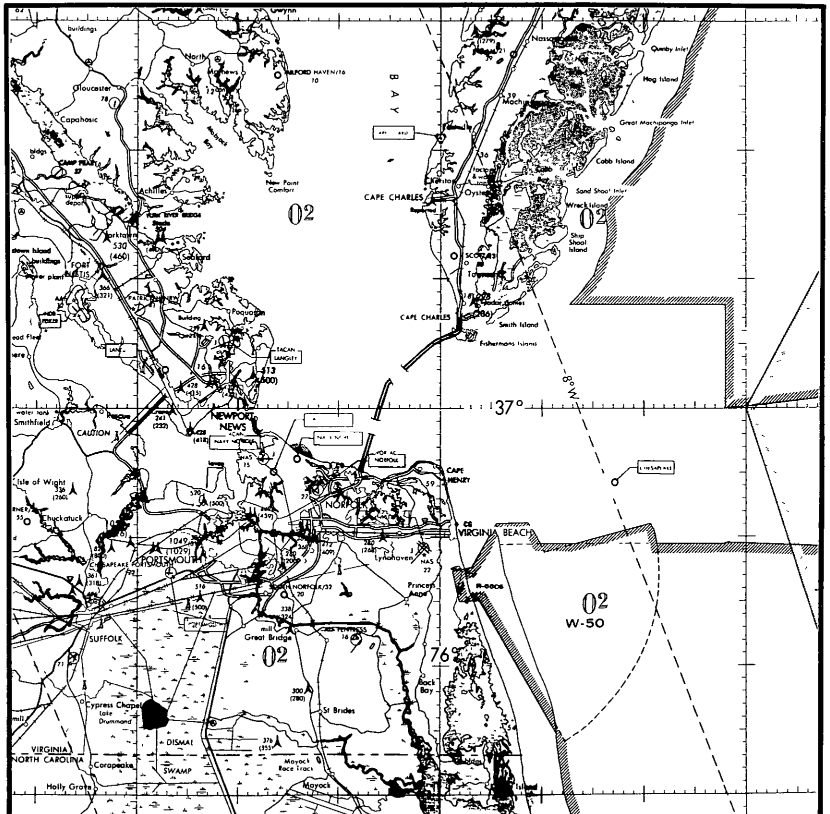


Figure 10

EXPERIENCE IN IMAGE PROCESSING

- DIGITIZING BATHYMETRY CHARTS
- DIGITIZING BLUEPRINTS
- PARTICLE COUNTING
- FINGERPRINT READING

Figure 1

Figure 2 is a picture of our image processing facility for digitizing the bathymetry charts. This digitizing program started 3 yr ago this December, and production started a year ago last July. So far 350 charts have been digitized and there are several hundred more.

The Navy digitizing system consists of a set of computer programs that run on a PDP-9 minicomputer (right side of fig. 2) The input is provided by a flying-spot scanner (background of fig. 2). The maps are photographed onto 70 mm film which is placed on a film transport inside the scanner. The digitizing is done frame by frame. The system was tailored to meet the needs of the Navy. Every point on the depth contours is not digitized, only selected points to form a grid of depths which is processed by a spline interpolation program to generate a uniform grid of depth data.

This system is an example of an application in which cartographic accuracy is not needed nor obtained. Sampling the depth information and then interpolating reduces the accuracy of the original data. Image processing techniques are used to automatically detect the location of a depth contour on the map and to follow it. The lines are followed not to encode them, but to carry forward the depth information from the last encoded point, thus eliminating the need for an operator to enter depth information again at the new point. This is an interactive system and the operator has considerable editing capability. This is a requirement for the system because the charts that we are working with, from the image processing point of view, are of poor quality; for example, frequently sections of lines will drop out when we scan the map image and transfer it to the computer for processing.

Figure 3 is a closeup of what the operator sees on the computer display. The grid overlaying the map detail is generated by the computer. The computer scans along each of these lines, called track lines, and processes the image data along the way. The processing is an algorithm that detects the intersections track lines and depth contours. What actually happens is that the computer follows the contour back to the previous track-line intersection and updates the depth for the newly detected point. When this operation is completed, it returns to processing the data along the track line until it encounters another intersection. If the computer is unable to follow the contour back to the previous intersection, the operator punches in the required depth information. The arrows shown on the display indicate that the intersections have been detected and processed, and the operator can thus verify that all the points were noted and add or delete information as needed.

The program for the Navy led to another program for processing map information that we call DIGIMAP. The technology developed for the Navy was extended to form a more general system for both digitizing or encoding map data and for postprocessing the information. Figure 4 shows the major hardware elements of the system: a flatbed image scanner (the input device), a minicomputer, and an interactive terminal. In the diagram, the blocks on the left form the digitizing system, and the blocks on the right form the postprocessing system.

Figure 2.--Calspan's SEASCAPE digitizing facility.

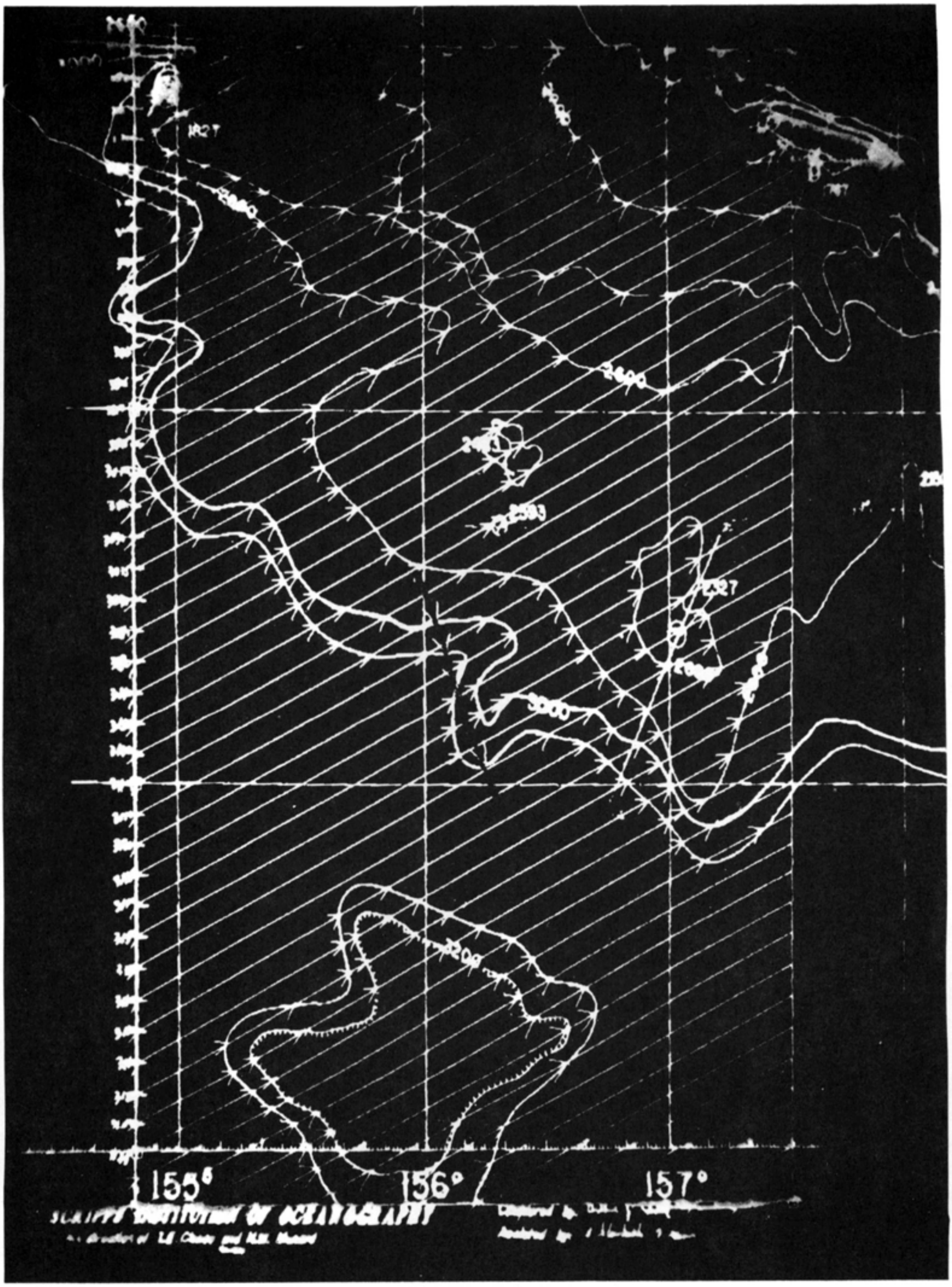


Figure 3.--SEASCAPE operator display.

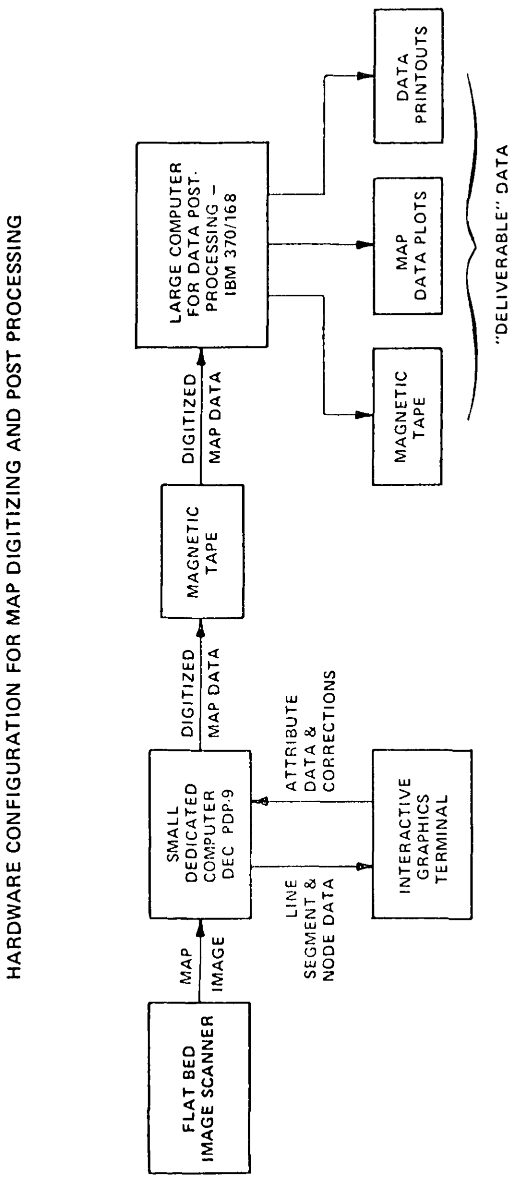

HARDWARE CONFIGURATION FOR MAP DIGITIZING AND POST PROCESSING
FLAT BED IMAGE SCANNER
MAP IMAGE
SMALL DEDICATED COMPUTER DEC PDP-9
LINE SEGMENT & NODE DATA
ATTRIBUTE DATA & CORRECTIONS
INTERACTIVE GRAPHICS TERMINAL
DIGITIZED MAP DATA
MAGNETIC TAPE
DIGITIZED MAP DATA
LARGE COMPUTER FOR DATA POST-PROCESSING — IBM 370/168
MAGNETIC TAPE
MAP DATA PLOTS
DATA PRINTOUTS
"DELIVERABLE" DATA

Figure 4

Figure 5 is another picture of our facility showing the flatbed image scanner in the forground (closeup in fig. 6). This device was developed in-house for scanning large documents. The key element of the scanner is a linear diode array which was discussed in an earlier presentation. The array has 512 elements and is mechanically driven across the table by x-y drive motors. Thus 512 points are scanned electronically. The scanner is a peripheral of the PDP-9 computer.

Figure 7 zooms in on the scanning head. The electronics are encased in the box; there is a simple optical system with a high-intensity tungsten halogen light. The room lights were turned off for this picture, but the scanner is normally operated with lights on.

I quite agree with Ray Boyle's earlier comment about it being difficult to separate digitizing from editing. Figure 8 illustrates the interactive editing operation with the DIGIMAP digitizing system. The scanning head is moved in a raster mode across the map being digitized, and the image data are transferred to the computer. In the computer, line-following image-processing algorithms locate, follow, and encode the lines on the map. Although the initial scan is done in a raster mode, the output is lineally encoded data; thus the system is a lineal or vector encoding digitizer. The map is processed in sections which correspond to half the size of the display. Note the bottom part of the display: the particular section of the map being processed is displayed, which is the raw image data as transferred to the computer. The digitized version of the same map section is displayed on the top of the split-screen. The operator compares the raw image data with the processed image data and then makes any needed corrections. There are a number of automatic editing functions, similar to the ones that Barry Moritz has described, which process the data before displaying. The key element in the minicomputer system is the disk that is used extensively to process the data.

Figure 9 is one example of a census tract map of Niagara Falls, N. Y. Our system has not been designed to process a map exactly like this one, but it is designed to process the unannotated version. We can process an annotated map like this, but it requires a lot of editing which runs up the cost of digitizing. The system is not designed to automatically edit out all of the alphanumeric symbols, although it does edit out some of them. Figure 10 shows the split-screen display of a section of the census tract map. This particular map was automatically divided into three sections for processing during the digitizing operation. Here, small boxes indicate nodes or intersections.

Figure 11 is another sample of a census tract map section which had been digitized and plotted with the tract code for each polygon. Here, only the polygons which are completely enclosed have been plotted. The plotting is done offline to the digitizing system; a CalComp-1136 plotter runs on our central IBM 370/168 computer facility. The maps are plotted to verify the digitized data. Figure 12 shows a tax map plotted from digital data.

In summary, Calspan Corp. has developed a semiautomatic digitizing or map-encoding system based on image processing technology. The system provides a means of encoding map data that is more accurate and less expensive than the basically manual systems which have been the major source of encoded data up to this time.

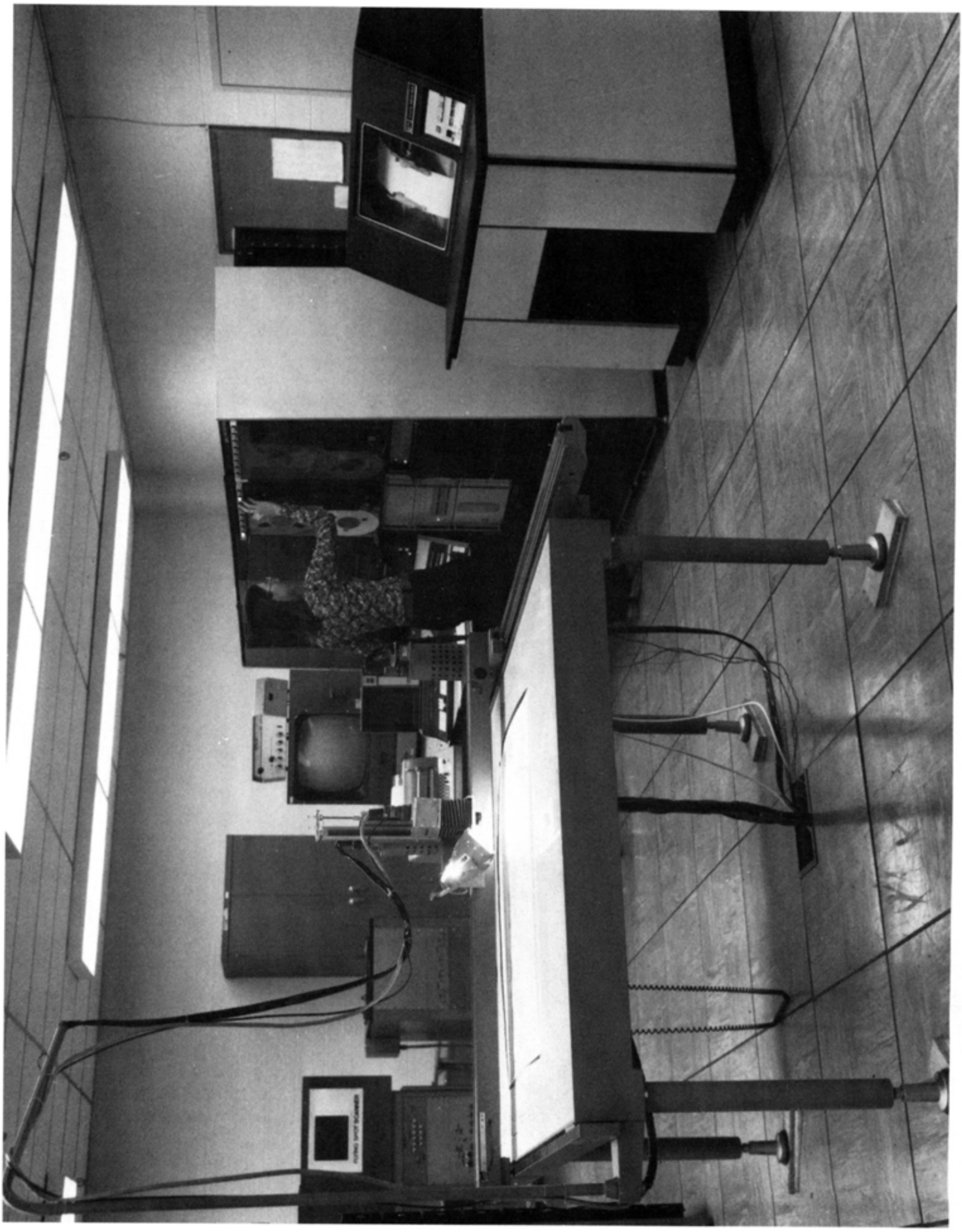


Figure 5.--Image processing laboratory.

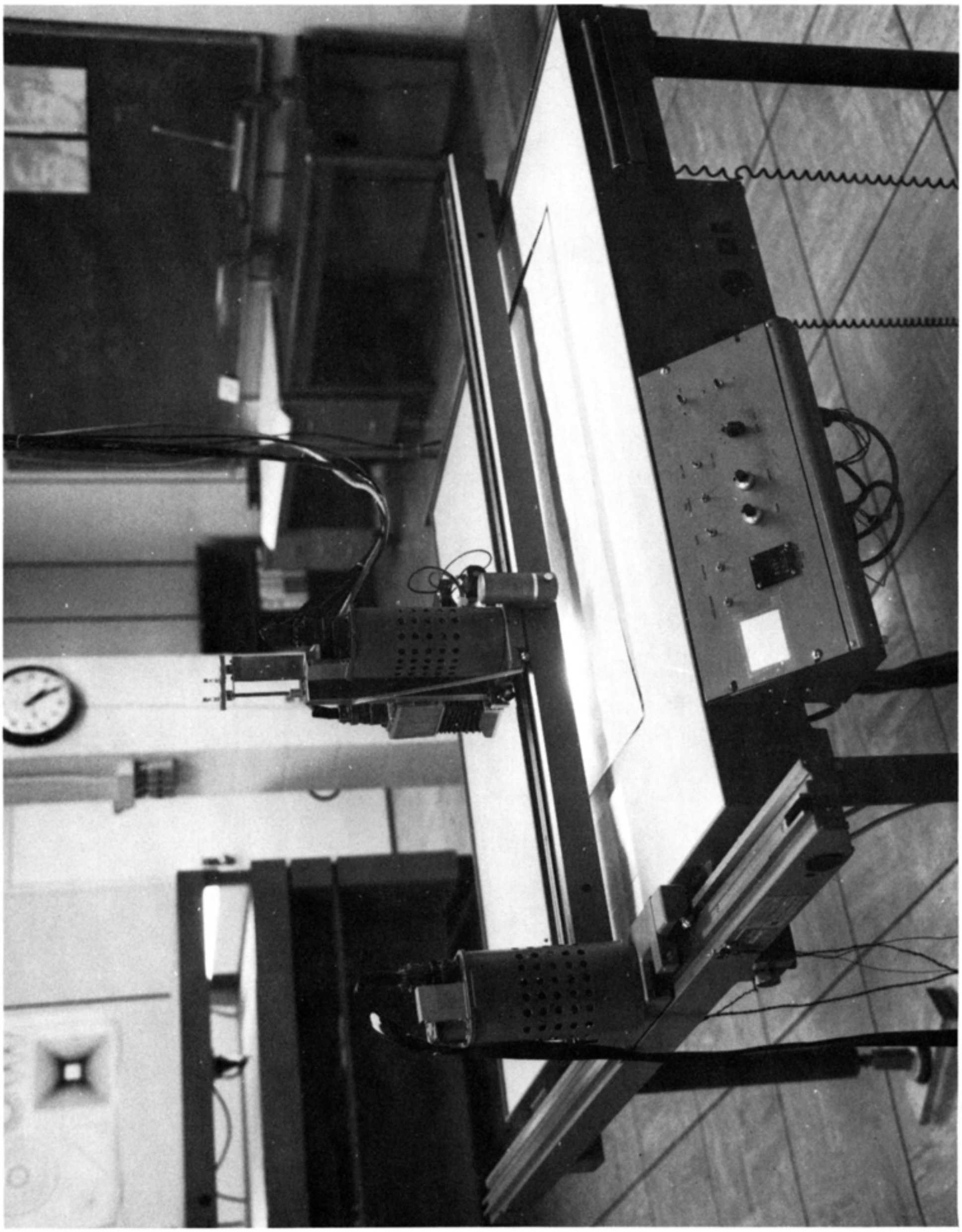


Figure 6.--Flatbed image scanner.

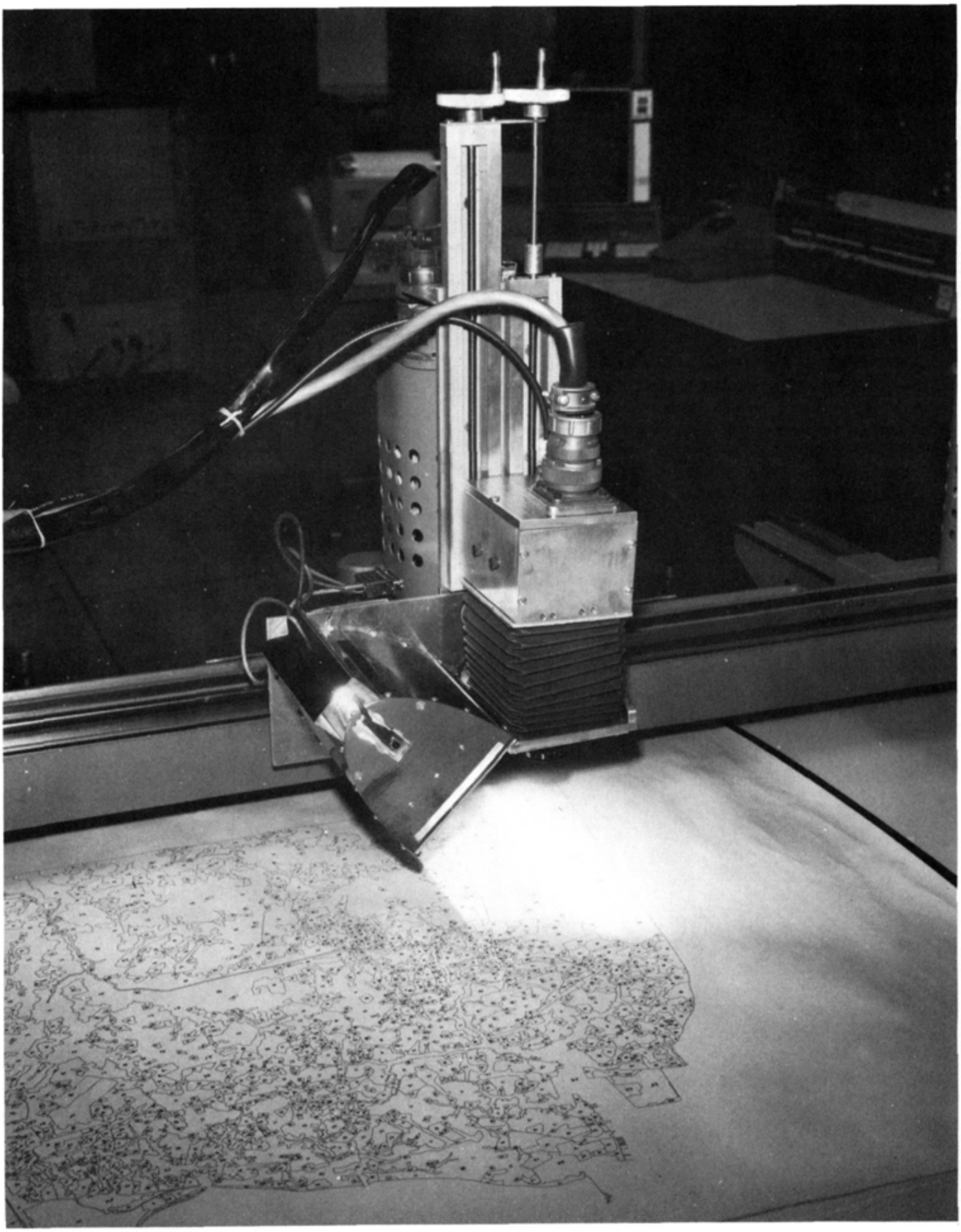


Figure 7.--Scanning-head assembly.

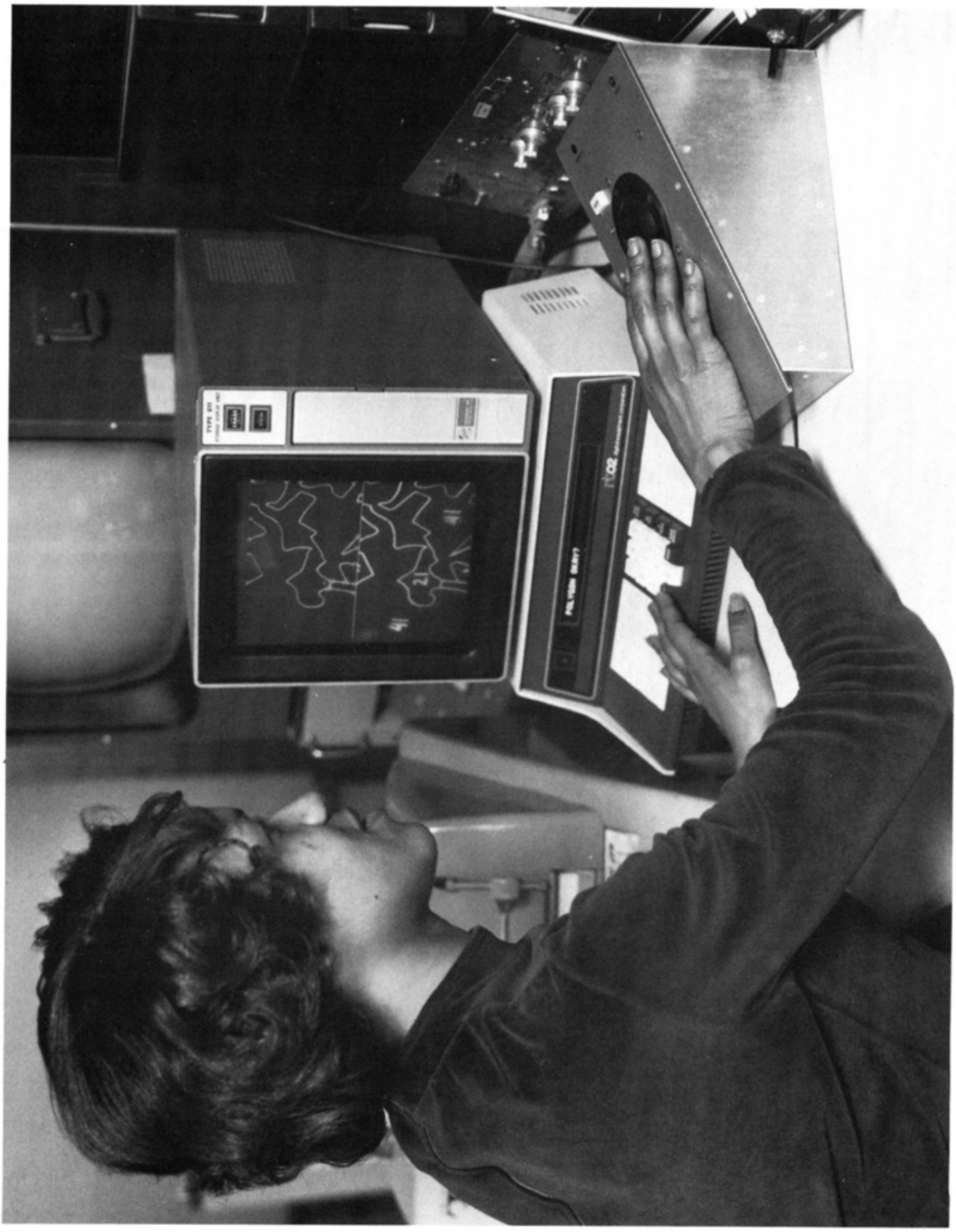


Figure 8.--DIGIMAP operator's split-screen display console.

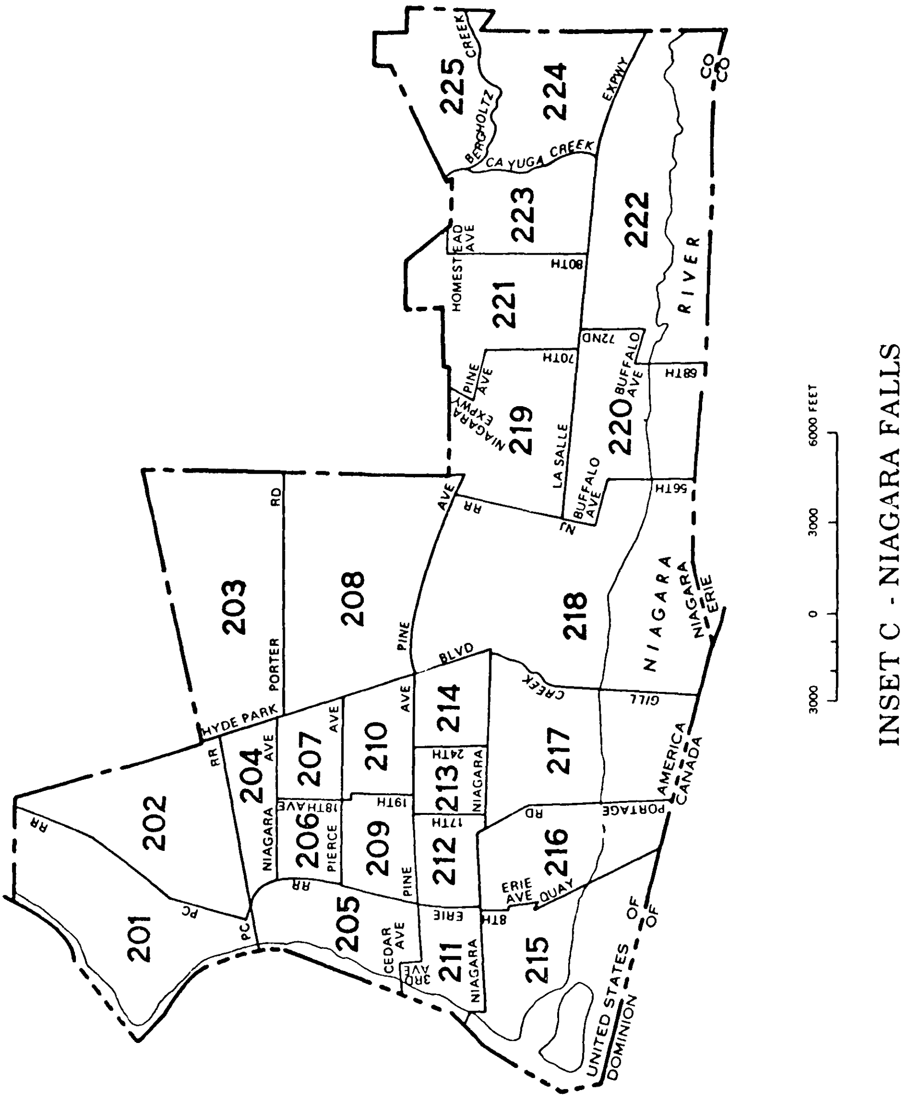


Figure 9.--Census tract map.

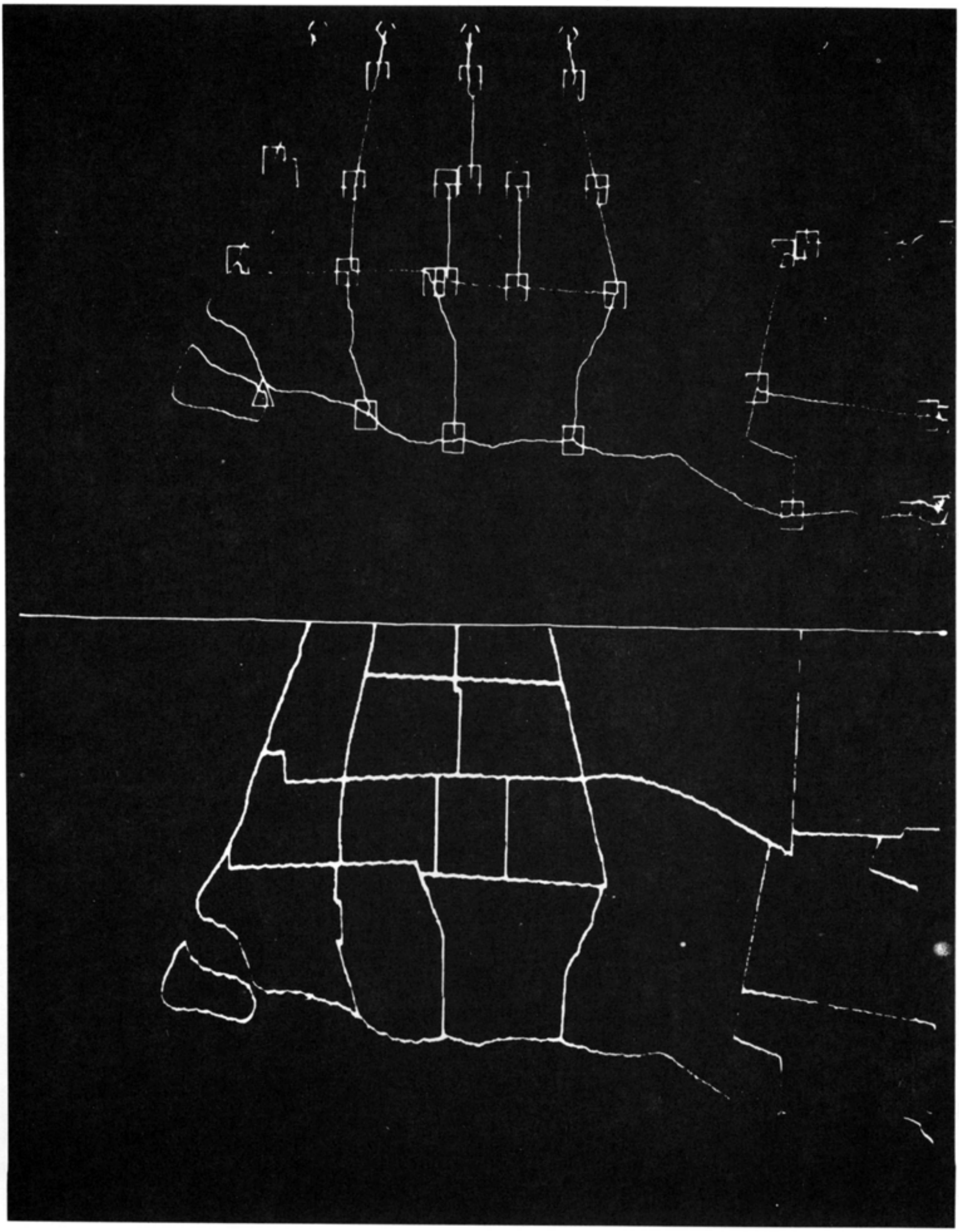


Figure 10.--Split-screen display of census map.

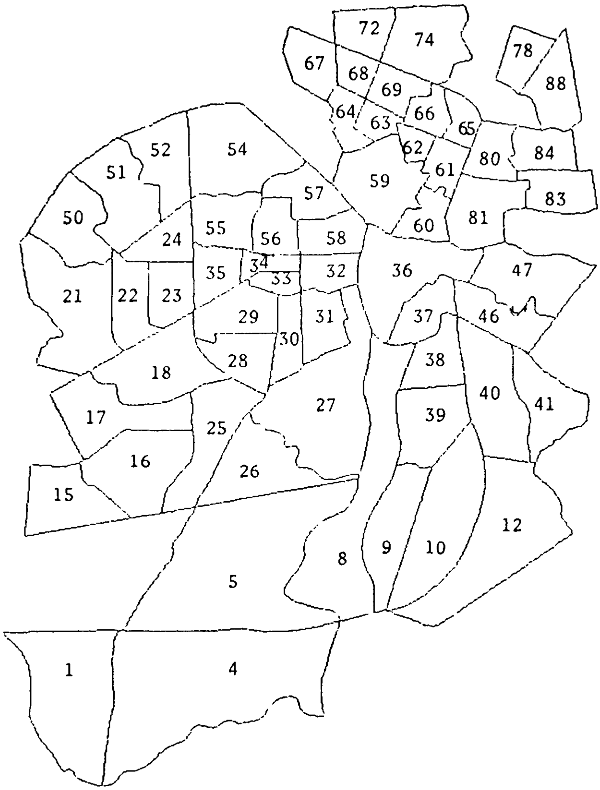


Figure 11.--Plot of digitized census map.

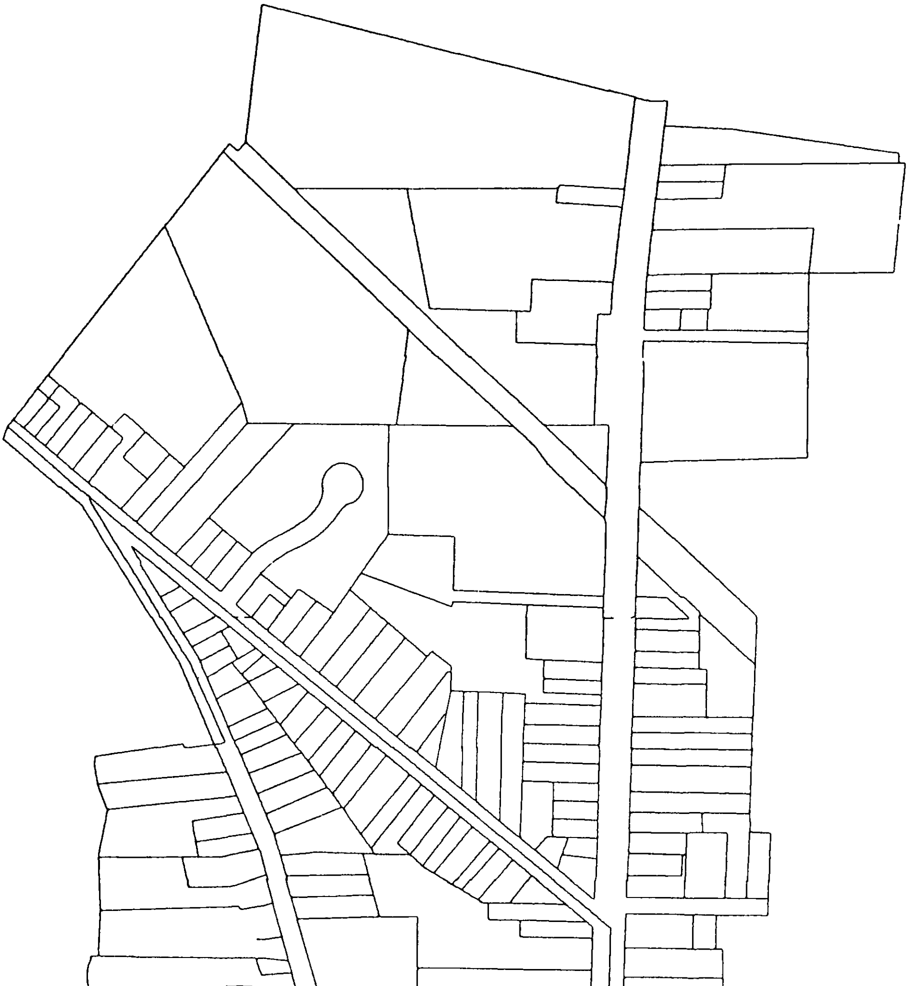


Figure 12.--Plot of tax map.

McDonough: As Olin Bockes pointed out, the question of data input has been around for a long time. The first Concord floating-arm graphic digitizer was built around 1960, and the system of 10 instruments that DMA Topographic Center has been using since then really generated the first bulk data file of graphic information--particularly terrain data--in the world. Surprisingly, one of the first large customers for that data was not a cartographic application but an electronic analysis application. The idea was that, in determining the interactions between radar sets and other electromagnetic devices, it is necessary to know something about the terrain where the devices lie. That was the first application, that I know of, for a large bulk file of digitized terrain information.

Those devices had a couple of elements that are extremely important in digitizing a very large amount of data. The purpose of the devices was to pick up terrain elevations from contour sheets. A lot of sheets were available; a lot of information was on them. One of the problems with digitizing contour sheets is that, when about 90-percent done, it is extremely difficult to review the digitizing and spot the last 10 percent of the contours. DMATC came up with a rather neat solution. They use the etched-zinc plate, of somewhat maligned reputation, and simply fill the lines with a colored water ink; as the digitizer is moved along the lines, the ink is scratched out. Thus, where the ink was scratched out, the line was digitized, and vice versa.

Lately, some other techniques have become available which have the same general characteristics; they rely on raster scanning the source material as opposed to linear detection. The raster scan approach has several interesting aspects. For a given-size sheet, raster scanning takes the same amount of time, regardless of the density of lines; a very dense contour sheet can be digitized in the same length of time as a sheet with one or two shorelines. Therefore, it is an extremely attractive technique to use when there is a large amount of information on a sheet. For certain applications, the raster form inherently contains a level of information separation or featurization when the source material is basically color-separated.

Raster scanning and raster plotting (the inverse process) permit the digitization and reconstruction of the graphics in a form that can be nondeteriorating, in contrast to the product of a series of photographic operations. Digital data allow line widths to be restandardized every time that you plot; you can then store the information in fundamental form as a graphic. This is particularly of interest to people who use perhaps one or two maps per application--utility maps, certain city planning maps, and similar materials--where the information is more readily stored in the graphic form, but a certain amount of editing or updating is desirable in digital form with an interactive editing system.

One difficulty with the raster technique is that featurization information is very crude. There is essentially no identification of individual linear features with a few exceptions, most notably the contour sheet. A couple of customers are successfully digitizing contour sheets entirely automatically--not for cartographic purposes primarily, but for terrain representation. Another customer has developed techniques for raster scanning well-prepared city planning charts, processing the data through a very large computer system, and producing lists of blocks, block sizes, street locations, associations of street-number ranges with certain block identification, and census tracts.

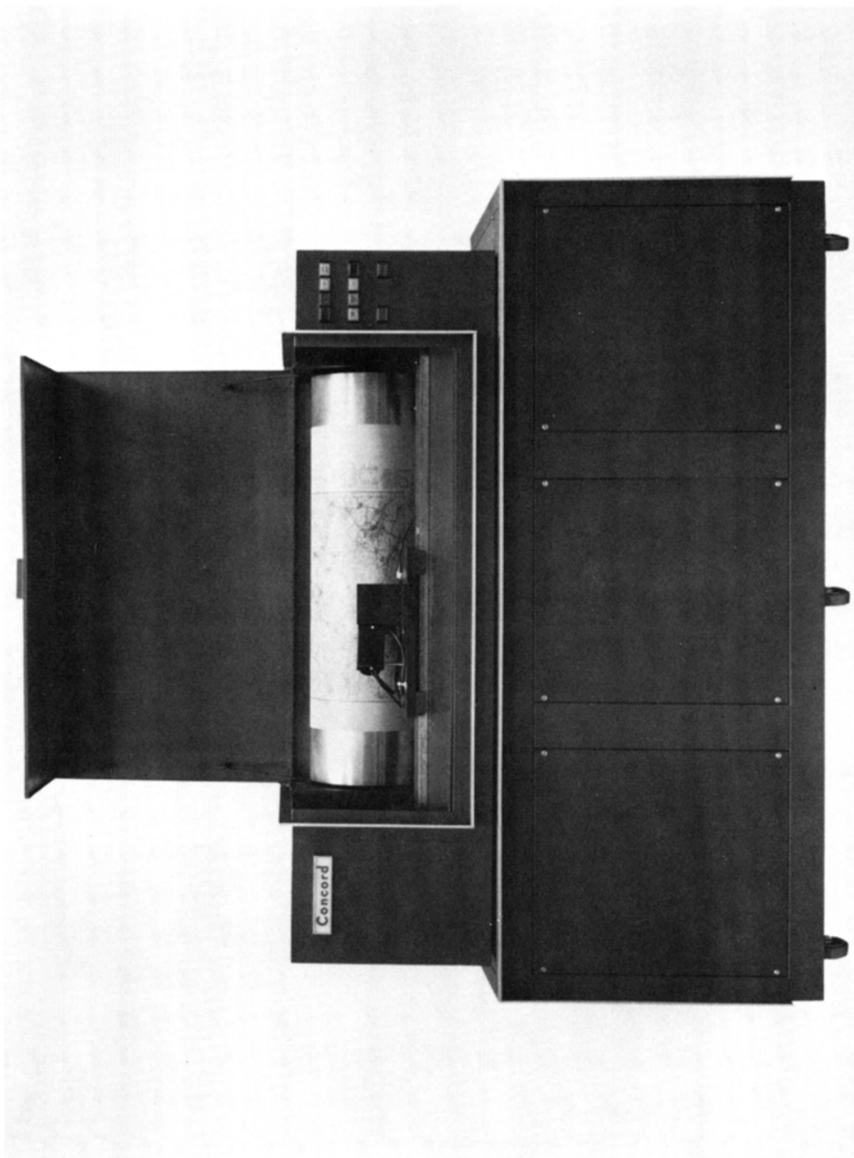
To go further requires some conversion of the raster data to linear format. The simplest way to do that is to turn a line-tracking algorithm loose on the raster data, let it track along the line until it comes to an intersection, and call each line segment between intersections a separate microfeature. Then a person will edit or connect the microfeatures representing one macrofeature. Each river, each road, and each contour on a composite sheet would be separated into little pieces which someone has to hook back together again.

In the last couple of years, a combined system was developed for RADC in which raster data are processed in connection with a manual digitizer. The manual digitizer works with a second copy of the source document. By rapidly digitizing approximately over a selected feature, the operator triggers the system to search the raster data for the specific location of lines most closely fitting the rough track data. The particular device now at RADC uses an online raster system in which a second copy of the map is on a scanner and as the operator makes a rough track, the scanner actually does the detailed digitization. The same sort of thing can be done by accessing data that were previously in a raster file.

Raster data appear to be much more bulky than linear featurized data. On the other hand, for materials such as dense contour sheets, the additional bulk is not spectacular if reasonable data compression techniques are used on the raster plotter. For various reasons many of the practical data base file systems that are based on linear data use a very high density of data along the linear feature. One of the reasons is to provide an easy means of aiming at a feature and finding a point. As a result, a reasonably sized raster file requires two or three reels of raster tape, as does roughly the same amount of information in one of the more densely recorded types of lineal format.

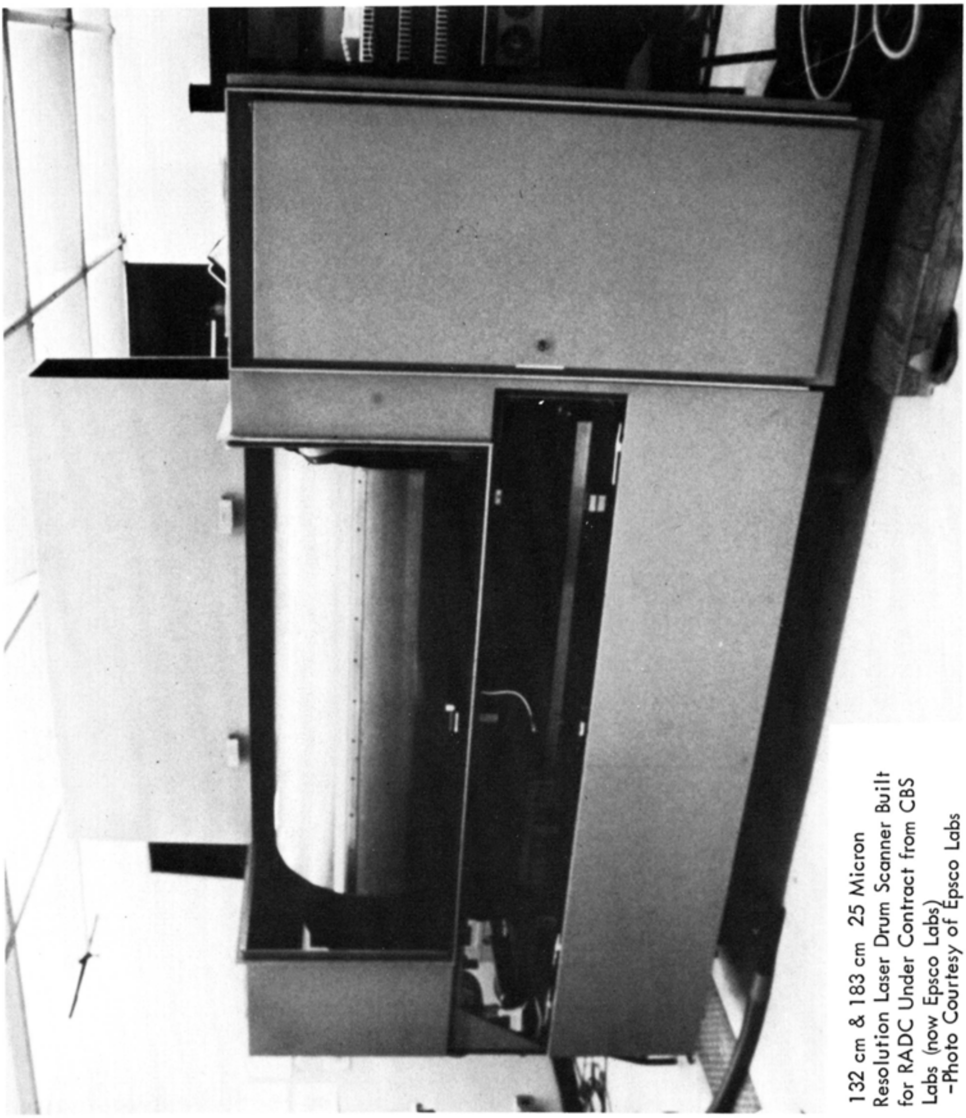
It is also possible to edit raster data directly in a raster format. With an interactive editing system it is possible to build up a small section of corrected data by any of the techniques--linear, raster, or whatever--and simply cut or "window" that data into the basic file. RADC has developed techniques for cutting in edited data and making the lines fit.

Concord Control has built a couple of raster scanning systems (fig. 1) that are being used in Europe for generating terrain elevation information. Presently Concord is working as a subcontractor for CBS Laboratories in the development of two very large, very high-speed raster scanner plotters for RADC (to be installed at DMATC and DMAAC). The systems measure about 52 by 72 in and are capable of producing either a scan file or plot in 30 min on a full-size sheet of film. It is a drum type scanner (fig. 2); and the scan carriage scans six tracks simultaneously around the drum, advances six lines, scans the next six tracks, and so on. Basic resolution is 25 μm, or 1,000 lines/in, for both axes. Concord also has developed, again mostly for RADC, a series of interactive graphic systems which combine linear plotting and linear digitizing capability.



Concord 90 cm x 81.92 cm Drum Raster Digitizer - Mark 3

Figure 1



132 cm & 183 cm 25 Micron Resolution Laser Drum Scanner Built for RADC Under Contract from CBS Labs (now Epsco Labs)
–Photo Courtesy of Epsco Labs

Figure 2

The old days of the operator laboriously following lines with a cursor or pushing a phonograph needle along grooves in an etched-zinc plate are over; digitization of a map file probably requires a little more human effort than went into the generation of the file on the first place. Now there are very rapid techniques for getting data into a digital data base by means of raster scanning. Some very sophisticated and reasonably efficient techniques exist that will convert raster data into editable form, by working in subsections of raster data or by derasterizing the data into linear format. One can take a composite raster file and extract specific isolated features desired for compiling a new product.

Peucker (Simon Fraser Univ.): When you raster scan an image which has a lot of dirty spots--ink and so on--you catch that, and you have to use quite an elaborate system and be reasonably efficient to get it out. With automated line following, you have a similar problem with this cutting across or not finding links. Therefore I really would like to see a description of a think-automated digitization process where difficult corners and difficult spots are marked beforehand by an operator and eliminated during the process of scanning rather than in an empty process.

McDonough: My own approach to the question of human interaction is that you should do as little preprocessing for the machine as possible. I think it is a difficult and time-consuming job to try to outguess a machine. Being somewhat of a software-man myself, I have spent a lot of time trying to outguess machines. The opposite approach is to use a process in which you can take advantage of the characteristics of the data and just ignore the dirt. The isolated dirt becomes a small feature whose total geographic dimensions are too small to accept. Similar criteria are used to distinguish isolated alphanumerics or other symbology from the linework.

On the contour sheet the lines are alledgedly all standardized in width. The system can select out of a glob of ink a standard-width line that either runs along the edge of the glob or through the middle of it; likewise contours can be peeled off one at a time from a coalesced area. These things have all been done. The difficulty is when a composite situation arises, like alphanumerics running into or across linework. In many cases, you don't have to clean all of that stuff out of the file unless you are looking for a total file of clean data. Such things as locations of telephone poles and management information can be extracted without cleaning the file.

The city planning system was a large IBM 750; the DMA system uses 1108 processing with a complex of minicomputers. Several hours/sheet are required, but there are only 35 sheets, so it is not a ridiculous application.

Bockes: I can't resist the suggestion that part of the problem raised by Peucker about ink spots and such may be caused by using second-generation information which is already in two-dimensional map form. We should not forget that terrain is three-dimensional information that might be better collected before you get to the map stage. This third dimension is an important factor; we may have damaged our digitizing by putting it first in a two dimensional map form and then trying to get it back into three-dimensional domain.

Wohlmut: Although i/o Metrics is concerned with all aspects of information retrieval from film, I will discuss only the method by which we digitize film-based data with the system that I outlined quickly in a previous session. As a service organization, i/o Metrics caters to a number of users with different kinds of data. A good part of the digitizing does not involve a direct correspondence between input film data and x-y coordinate output. For example, in high-energy physics we deal with geometric coordinates on a piece of film and desire output as kinematic variables of energy and momentum. Another project involved scanning 20 ft of coiled wire and producing a histogram of the variation in thickness.

We have attempted to set up a system that produces the final digital output in as easy and general a fashion as possible. In some applications the approach is raster scanning or a modified raster scanning with strings of density information. However, automatic line-following is used in most applications because the total amount of processing and editing time is minimized. We use a laser scanning device which has x-y random position capability and can sample or generate position, direction, density, and line-width information under computer control. Currently, we have three devices: one is in the final stage of development (should be available in 6 months) and will be able to scan a 5-ft-square area: the other two will scan a 105-mm-square area with an absolute accuracy of 2 μm. The systems enable sampling the x-y position of a point, line, or other physical attribute at any time as well as instructing the probe to move to certain x-y coordinates.

Besides knowing x-y coordinates at all times, two other variables are sampled: the percent transmission of light through the information base and, for automatic line following, the pulse width of the object of interest. The system operates with FORTRAN programing which is highly modular. At any time, if one can either precisely or statistically define a given object of interest in terms of the above parameters, the machine will very satisfactorily decide whether to try to automatically follow a line one way or another, whether to ignore a piece of information, whether to accept it and continue, or whether to seek operator assistance.

The system (fig. 1) basically consists of three elements--the film-plane digitizing stage, an interface to a computer (currently a Unichannel PDP15), and a control console which permits an operator to monitor the operation and to interact with the device and/or the computer at will. The latter feature enables on-line editing and correction and correction of unprogramed conditions.

The characteristics of the scanner itself include those of the deflection system that I described very quickly in an earlier session. Figure 2 shows the optical paths of the system. The film plane is set 1 m from the deflection system, largely for our own convenience since most of the formats that we have dealt with were under 105-mm on a side. The deflection mirrors move over the film plane under computer control, and the probe beam, a 1.5-mw helium neon laser, is focused down on the film. The beam passes through some beam-collection optics, through the film, and into a photomultiplier (PM) tube which is clocked in with the other information that I described previously, so that we know at any given time where we are and what the signal looks like.

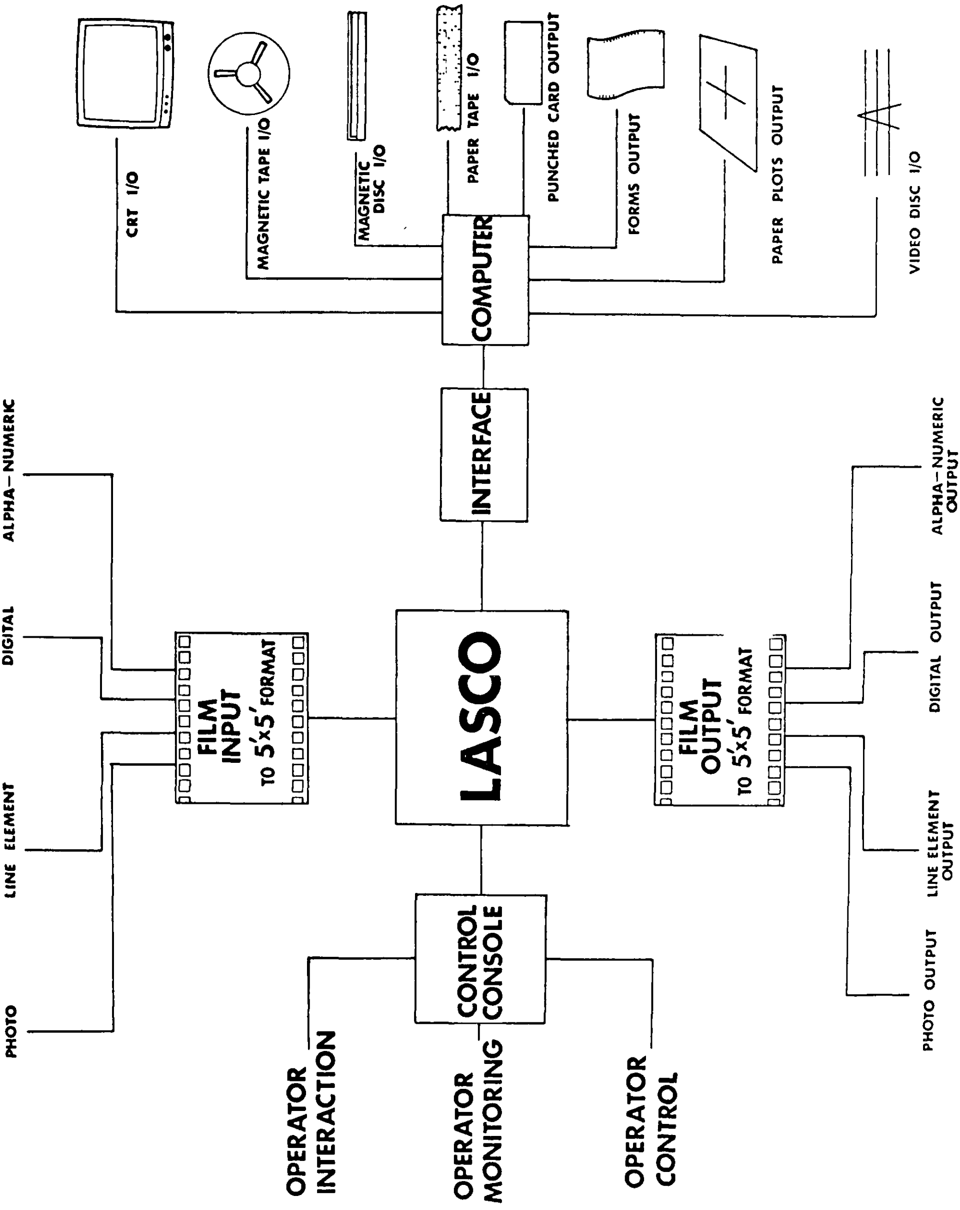

CRT I/O
MAGNETIC TAPE I/O
MAGNETIC DISC I/O
PAPER TAPE I/O
PUNCHED CARD OUTPUT
FORMS OUTPUT
PAPER PLOTS OUTPUT
VIDEO DISC I/O
COMPUTER
INTERFACE
ALPHA-NUMERIC
DIGITAL
LINE ELEMENT
PHOTO
FILM INPUT TO 5'x5' FORMAT
LASCO
FILM OUTPUT TO 5'x5' FORMAT
ALPHA-NUMERIC OUTPUT
DIGITAL OUTPUT
LINE ELEMENT OUTPUT
PHOTO OUTPUT
CONTROL CONSOLE
OPERATOR INTERACTION
OPERATOR MONITORING
OPERATOR CONTROL

Figure 1

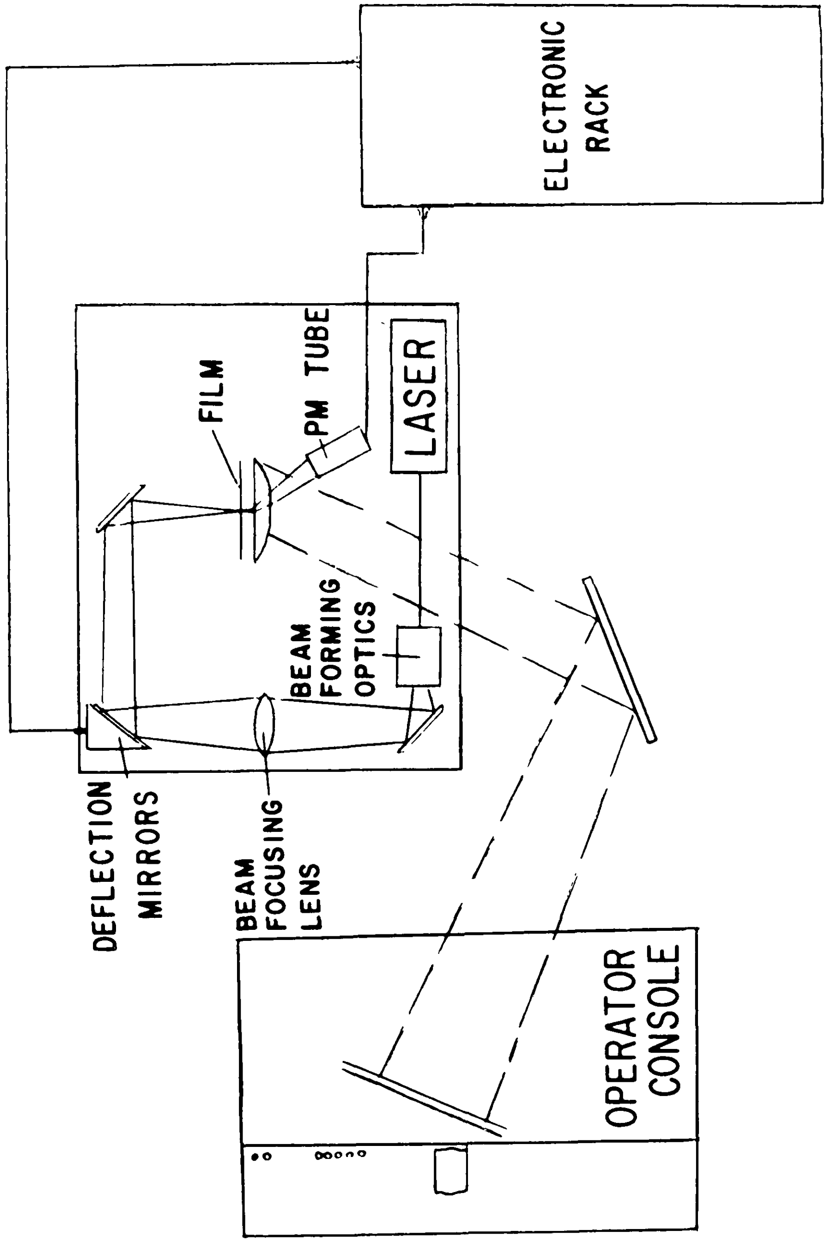

ELECTRONIC
RACK
FILM
PM TUBE
LASER
BEAM
FORMING
OPTICS
DEFLECTION
MIRRORS
BEAM
FOCUSING
LENS
OPERATOR
CONSOLE

Figure 2

To assist the operator in monitoring, a reverse beam of mercury arc light goes back through the system and produces a large screen blowup of what is being digitized. At the same time, a closed-circuit TV monitor gives a picture of the immediate area around the probe. The console also has a Tektronix 611 display which graphically shows the actual coordinates digitized. There is a teletype for commanding special functions. In addition, there is an oscilloscope which shows the signal quality as the scanner samples the film.

Raster scanning is performed in a standard fashion. A 10-µm spot can be produced with a 1-m throw. The spot can be moved randomly or sequentially in x and y.

For automatic line-following, a 10-µm spot rotates in an orbit with an arbitrary radius up to about 2.5 mm. The radius is arbitrary to the extent that it can be programed or it can be set manually; however, there are particular radii that are known to work best for certain applications. The computer knows the precise center of the orbit to an absolute accuracy of 2 µm. The radius is determined fairly precisely by a calibration technique we use; for a 1-mm radius, the accuracy is again about 2 µm. The angular position of the probe is known at any given time to an accuracy of 1 part in 4096--a somewhat arbitrary limitation because accuracy is only restricted by the size of the counter and the rate that the image is spinning. For most line-following applications, the point on the circle spins at the rate of 150 r/sec, which is a little faster than the software will handle most of the decisionmaking. However, one is able to automatically line follow and digitize on any arbitrary curve between 50 and 100 points at a spacing anywhere from 1 µm to 2 mm. The choice of coordinate point spacing depends on the output desired and partially on the complexity of the curves. For rapidly varying curves, the sampling rate is high in order to preserve the integrity of the data and to facilitate searching for the next data point under computer control. Digitizing is typically carried out at 5- to 10-mil increments at about 100 points/sec.

Part of the programing for line following involves knowing the x-y coordinate location and knowing something about the history of the line, or track, that we are digitizing. As the probe crosses the track, the density of the track is measured by the transmission of light, and the track width is measured by the number of angular counts for which the measured density exists during a sweep of the probe. The PM tube displays a straight line when no data are present; when the probe crosses a track, the PM tube displays a dip or "bucket" of a width that is a function of the track width and of a depth that is a function of the density. The probe steps along the track while the computer records and predicts where to look next. The computer can be selective in sampling a particular angle; in other words, you have to sample density all the way around the probe orbit. As a history is built up the opening angle is selectively narrowed down to reduce any confusing effects of dirt on the sides or of tracks closely parallel to the track of interest.

The rotating point can be extended into a line segment. This was typically done early in the development of the digitizer for following bubble-chamber films, which really don't contain very well-defined tracks but consist of series of bubbles with sizable gaps

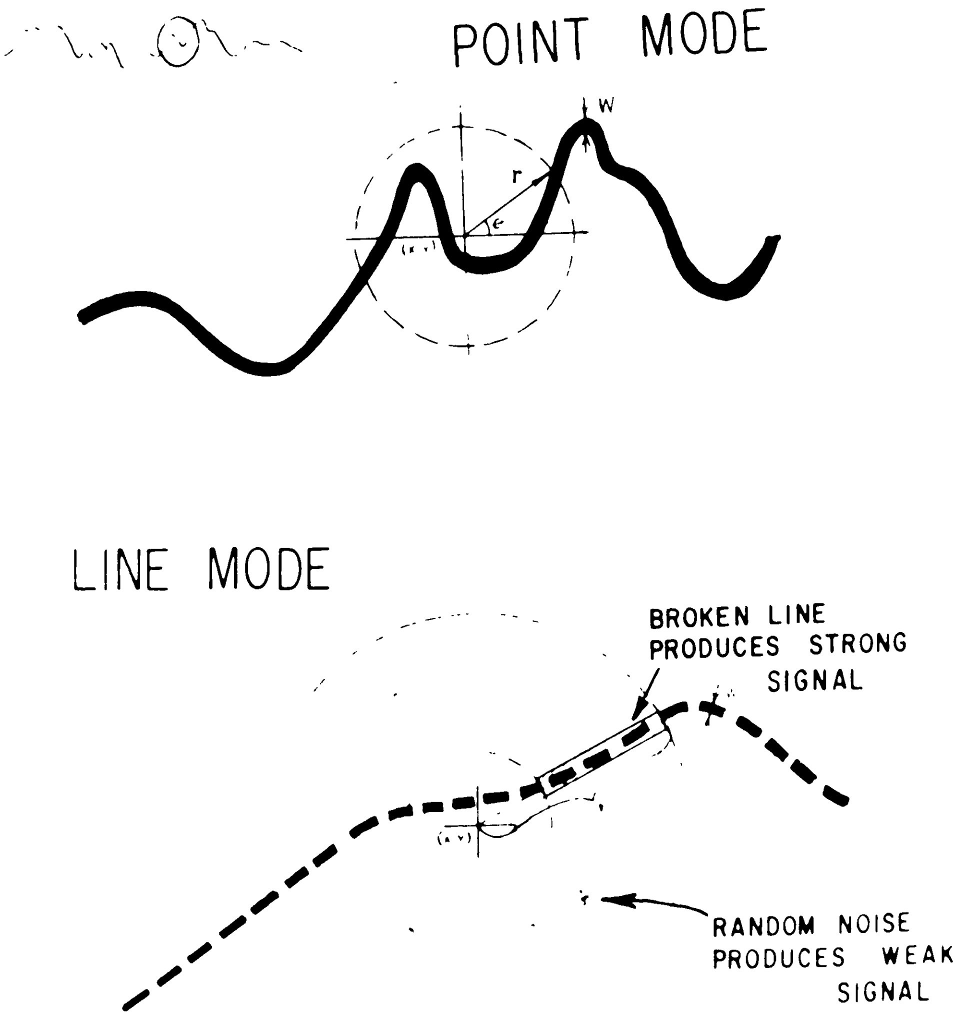

POINT MODE
W
r
LINE MODE
BROKEN LINE PRODUCES STRONG SIGNAL
RANDOM NOISE PRODUCES WEAK SIGNAL

Figure 3

on very low-quality film. Sampling (fig. 3) by averaging over an area larger than the 10-μm spot is not a very effective technique when you encounter tracks that rapidly change direction, but it is a most useful technique for "well-behaved" lines which have large gaps or irregular line width. In addition, the step size can be changed typically to handle drastic variations of the direction of the track. Effectively one can obtain only those coordinates of a line which will define the line variation with a predetermined accuracy.

Figure 4 shows a plot produced by the same probe used to measure the data by exposing the film under computer control rather than sampling. A standard land-use sheet is cut into 105-by 105-mm sections, digitized separately by using fiducial marks as common reference points, recombined, edited and perhaps digitized in selected areas, and then plotted.

The procedure for digitizing the land-use data shown in figure 3 was as follows: We were provided with a digital tape of a land-use data--label information on each of the polygons and coordinates of a point within each polygon. As the tape is read, the probe jumps automatically to the given center point in the polygon and then hunts in one direction for the start of a line element. The operator can override the procedure if the probe is not hunting efficiently and can position the probe manually. The probe automatically follows the track in either direction. When an intersection is reached, the probe "tells" the computer that a line segment was completed and/or a new line segment is starting, and then it proceeds to close the polygon. Thus the digitized data are coordinates of closed polygons with line segments which are coded appropriately; the output can be in any format that is compatible with the user's needs. This is all done with minimal editing. The plot in figure 4 represents 150,000 points at 10-mil spacing, which corresponds to 300,000 18-bit words. In this case, some areas were redigitized to see how well the data compared with the original measurements; after reconstruction the variance amounted to 0.001 in over a 5- by 5-ft area.

Figure 5 shows one corner of a map, the most extreme corner from the origin of the system. The darkest features represent the overlap between two adjacent segments. The dark lines are the remeasurements on top of the original measurements. The result is that we can combine data and achieve better than 1-mil accuracy over the entire area without manipulating the raw data other than translation and rotation.

In summary, we feel that we have demonstrated a viable technique for producing a digital data base which is accurate, efficient, and most effective and which is currently in operation.

Chrisman: I want the chance to question this in depth, but time just doesn't allow it.

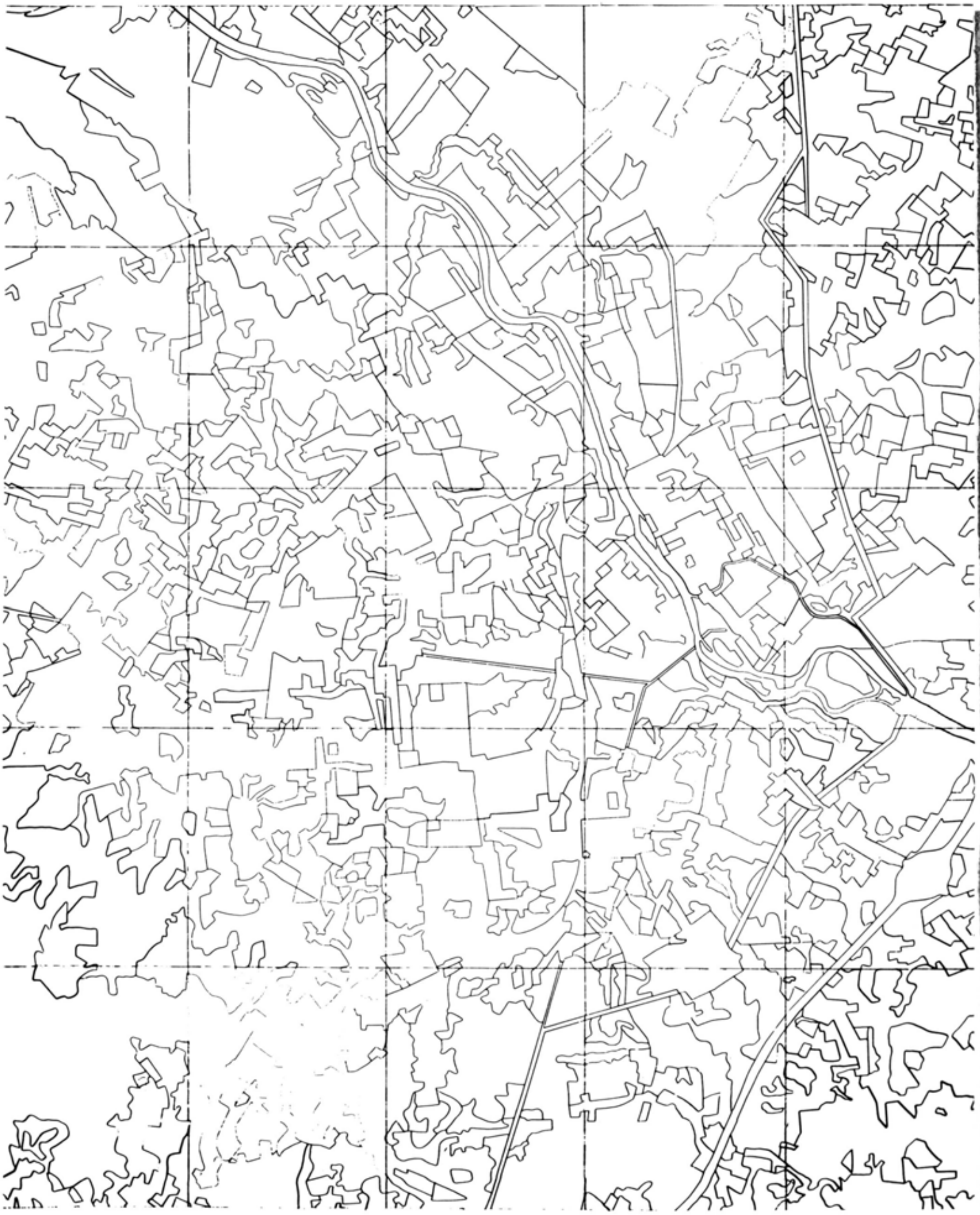


Figure 4

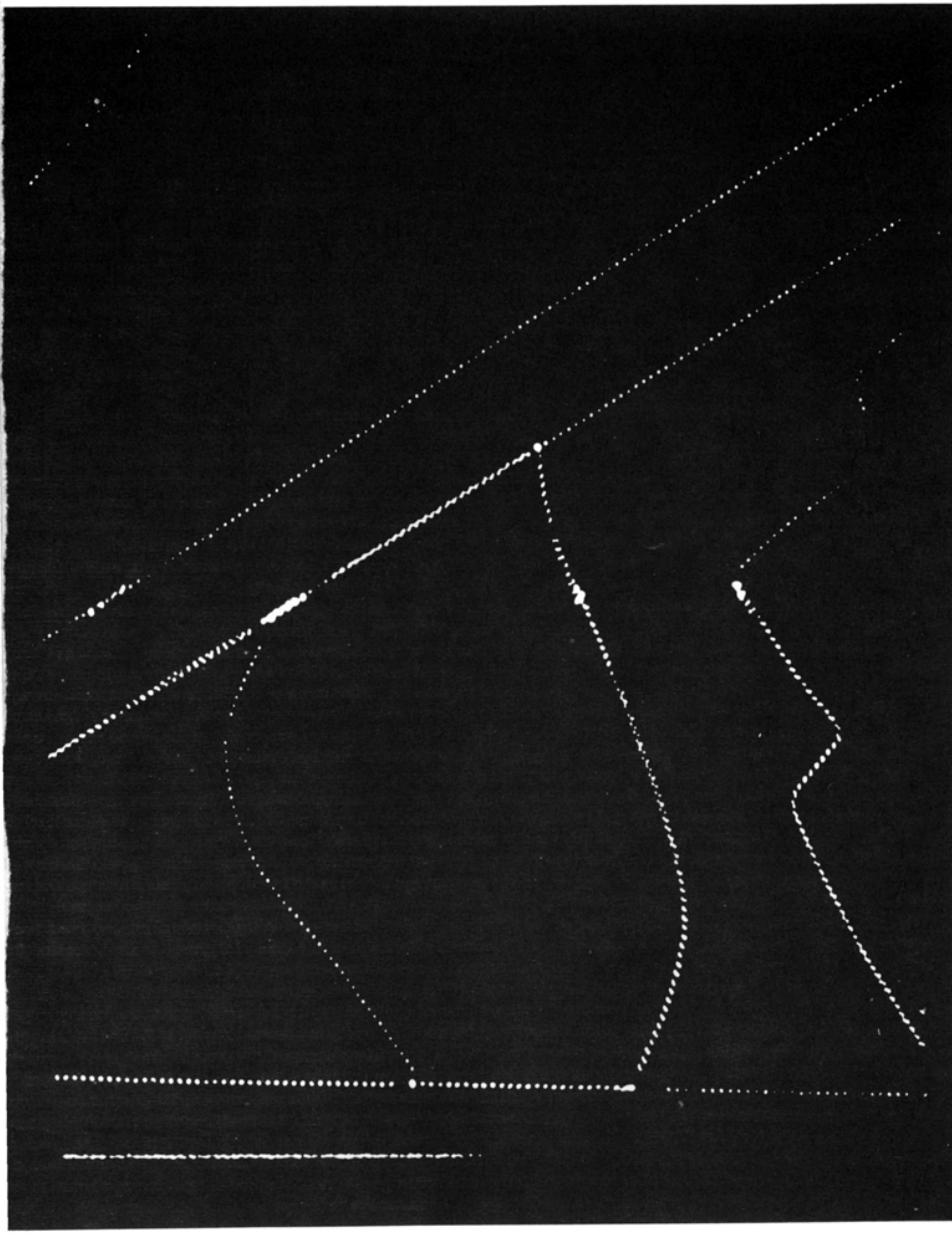


Figure 5

Stiefel (Analytical Systems Engineering): It seems that one of the threads running through the entire conference so far is that we have effectively buried manual digitization. Is it possible to digitize to National Map Accuracy Standards using manual digitization techniques, or are we stuck with automatic scanning which obviously means a larger investment? A lot of people here have a great deal invested in manual equipment. If it is true that manual digitization is not practical, where can we apply manual systems, and where must we go to automatic scanning?

Wohlmut: The land-use plot that I showed you, representing 150,000 points, was provided at a total cost of $450 for error-free output.

Stiefel: I have no way of determining your accounting method.

Wohlmut: That is all that it would cost you for an equivalent map. There are other things that would go into it. How would you use an automatic device completely? There is always some major operator intervention required; the more manual intervention, the higher the cost. I doubt that it is ever possible to have a completely automatic system. There will always be things that will require an operator's assistance.

Stiefel: Yes, that is one side of the question, but how about the manual-digitization people?

Chrisman: I don't see any point in a person recognizing the feature and sending the device to it. To digitize the DIME file by this system, the operator would be spending just as much time as he would with a manual system because he would have to identify each street segment, each address range, and the left and right polygon, which is precisely what he would have to do in the manual system. He would have to have previously entered the address range, the polygon centroids, and the polygon outlines.

You have a wonderous piece of hardware. I don't see how this hardware identifies geographical objects. I have seen what you have done for Lawrence Berkley Labs--a line scanner that has a real problem at line intersections and requires a lot of hand intervention. I saw that project; 70 percent of the effort was getting rid of the node errors and 30 percent was letting the machine digitize the rest.

Wohlmut: I can say categorically that such problems do not exist. You are talking about a pilot project that was conducted overtly, and the early stages involved quite a lot of development work. At this time, the machine recognizes corners and places in the corners, well within the accuracy requirements of the contracts.