

DEPARTMENT OF INTERIOR

U.S. GEOLOGICAL SURVEY

**GRIDVECTOR (version 1) : AN ARC/INFO AML PROGRAM TO EXTRACT
LINEAR FEATURES FROM A GRAY-SCALE IMAGE OF A PAPER
GEOLOGIC MAP**

BY

Feliks Persits

Open-File Report # OF97-713

This report is preliminary and has not been reviewed for conformity with U. S. Geological Survey editorial standards.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U. S. Government.

Although this program has been used by the U. S. Geological Survey, no warranty, expressed or implied, is made by the USGS as to the accuracy and functioning of the program and related program material, nor shall the fact of distribution constitute any such warranty, and no responsibility is assumed by the USGS in connection therewith.

U. S. Geological Survey, Denver, CO 80225

GRIDVECTOR, an ARC/INFO macro language program, provides a more effective method of accomplishing the first step of on-screen digitization of large and complicated geologic maps. The traditional method used is first, hands-on tracing of selected lines on a transparent mylar sheet. The mylar is then scanned in black and white mode and the scanned image converted to lines by the ARC/INFO GRIDLINE or ARCSCAN commands.

Disadvantages of the traditional method are:

1. Much more time is needed to prepare drawings on mylar when features on the original geologic map are very tiny and the number of polygons is large (e.g., hundreds or thousands).
2. Poor accuracy is common because of accidental shifts of the mylar while tracing lines and because lines may be inadvertently omitted.
3. Adding these omitted features requires rescanning the mylar and consequently introducing new spatial errors when matching the new features to the existing data.
4. Identifying the same line on the screen and on the original paper map is time consuming and tedious work.

GRIDVECTOR is designed to trace boundaries of geologic features from a gray-scale scanned image of a paper geologic map. The advantages of this method are:

1. A gray-scale scanned image is far less expensive than a color image and the image itself needs much less disk space.
2. Even at relatively low resolution (200 dots/inch), most geologic features are retained in the gray-scale image.
3. The color separates of the original color map, if available, can be used as input to GRIDVECTOR to trace colored boundaries.
4. Adding newly requested features or correcting existing ones is simple because the same gray scale image is used.

Scanned images are transformed into a grid, using the ARC/INFO IMAGEGRID command and georeferenced using the ARC/INFO REGISTER and RECTIFY commands. The georeferencing is done mainly by using latitude/longitude crosses or other features with known coordinates as references.

The output of GRIDVECTOR is a line and polygon ARC/INFO coverage. Here lines are supposed to be one of the following features on the gray-scale image of the original paper map:

- a) the centerlines of the dark, relatively thin, strips (polygon outlines, faults, etc. on the paper map);
- b) boundaries between areas with different gray levels (different color or pattern on the paper map).

The method presented in this report extracts lines from a gray-scale image. These lines (strips of pixels) can be easily seen on a computer screen in ARCEDIT, when the image has been used as the background. If the original map is of good quality (i.e., good line contrast and a small number of different patterns and text) the GRIDVECTOR program can extract 50-60% of the lines, saving a great amount of handwork and time. The rest of the digitization, and correction, can be done effectively along with the geologic attributing in ARCEDIT. This helps avoid errors caused by separation of these processes in time.

The program is based on the assumption that a dark (mainly black) line outlines every polygon (geologic feature) on a paper map. Such a line in gray-scale is represented by a thin strip of pixels with minimum gray values. Features that can be treated as noise are other dark lines, such as rivers, and text annotations. To reduce the number of lines and polygons created because of such noise, the average slope of gray-scale pseudo relief was calculated. The average gray levels are about the same on both sides of the dark line representing a river, text, or coordinate grid. By contrast, a geologic polygon gives a jump in gray level because of the different color or pattern on either side of the line. The user has a choice of selecting lines by the minimum of gray values, by average slope, or

by both methods. Before calculation of an average slope, several consecutive smoothing filters are applied.

The program is based entirely on standard ARC/INFO, ARCSAN and ARCEDIT commands. ARC/INFO's GRID program is limited to a maximum filter size of 7 cells (pixels). Therefore, the optimal image resolution is about 200 dots/inch, giving a width of 4-6 pixels for dark lines.

The program uses ARC/INFO's GRIDLINE command to create a line coverage; the final results are very dependent on the effectiveness of GRIDLINE.

An easy way to apply the GRIDVECTOR AML is to select parameters on a relatively small portion of the input grid. The user can then experiment as many times as necessary, because the program deletes all previously made files. The goal is to select parameters that allow an optimum output that satisfactorily filters out the noise without removing too many geological boundaries.

The following figures show some examples.

Figures 1-5 show the GRIDVECTOR application on a fragment of a 1:7,500,000 scale geologic map of relatively poor quality (i.e. pale colors, variable line width, very tiny polygons). Figure 1 shows the input grid with arrows pointing to latitude/longitude lines that are considered to be noise. Other noise features are text labels and a background dot pattern. The final line coverage (after GRIDLINE) in figure 4 shows advantages of the GRIDVECTOR algorithm. Coordinate lines and many of the text labels are completely filtered out. Some of the polygons look completed and are ready for geologic attributing. The results in the lower-left corner are of poor quality; further editing and correcting need to be done before geologic attributing can start because there are many small polygons and small distances between lines in comparison to the GRID filter size. At the same time, in figure 3 (binary image after grid processing which serves as the input for GRIDLINE) the white lines are more contiguous and more like the original. This illustrates the fact that the effectiveness depends on GRIDLINE. Figure 5 shows another run on the same grid, using different parameters; - a higher threshold for dark lines and average slope. There are fewer false lines in this coverage, but at the same time more

digitization will be needed. Figure 2 shows an intermediate average slope grid where the white pixels are maximums (arrows point to some). The coordinate lines and many of the text labels have disappeared but much correction and digitization is still needed. Half of the desired lines are present and therefore the same amount of handwork has been saved.

The second example shows another GRIDVECTOR application. The goal was to select and trace sea depth contour lines from a gray-scale image (Figure 6) in which areas between the contours are filled in by different patterns. In spite of this image noise, lines on the final coverage (Figure 7) follow the original contours accurately. Minor corrections, like filling in gaps and removing false lines, will still need to be done.

Conclusion:

GRIDVECTOR has been created as a helpful tool for paper map digitization, that continues to be a most tedious and time-consuming task in new GIS projects. The standard ARC/INFO GRIDLINE or ARCSCAN subsystem can be applied only to binary images and therefore requires handmade mylar copies of maps. GRIDVECTOR provides a more efficient gray-scale scanning method of moving a paper map into ARC/INFO format.

Figure captions

- FIG. 1** – A fragment of a 1:7,500,000 scale geologic map of relatively poor quality, i.e. pale colors, variable line width, tiny polygons. Gray-scale scanned image of this fragment has been converted to grid by IMAGEGRID command and used as input for GRIDVECTOR program. Arrows point to a latitude/longitude lines that are considered to be noise. Other noise features are text labels and background dot patterns.
- FIG. 2** – An intermediate average slope grid, from map fragment in figure 1, where the white pixels are maximums. Latitude/longitude grid lines and many of the text labels have disappeared.
- FIG. 3** – Binary grid, from map fragment in figure 1, after completion of input grid processing which serves as the input for GRIDLINE. Noise features including latitude/longitude grid lines are converted to groups of speckles that are in turn removed by the GRIDDESPECKLE command of ARC/INFO.
- FIG. 4** – Final line and polygon ARC/INFO coverage after GRIDLINE, from map fragment in figure 1. Coordinate lines and many of the text labels are completely filtered out. Some of the polygons look completed and are ready for geologic attributing. The results in the lower-left corner however are of poor quality and further editing and correcting needs to be done before geologic attributing can start.
- FIG. 5** – Another run on the same grid, as figure 1, using as different parameters: a higher threshold for dark lines and average slope. There are fewer false lines in this coverage, but at the same time more digitization is needed. Map fragment from figure 1.
- FIG. 6** – A fragment of a map at 1:7,500,000 scale showing sea depth contours that are to be extracted. Areas between the contours are filled in by different patterns.
- FIG. 7** – Final line ARC/INFO coverage of the map fragment, figure 6. Minor corrections, like filling in gap and removing false lines, will still need to be done.

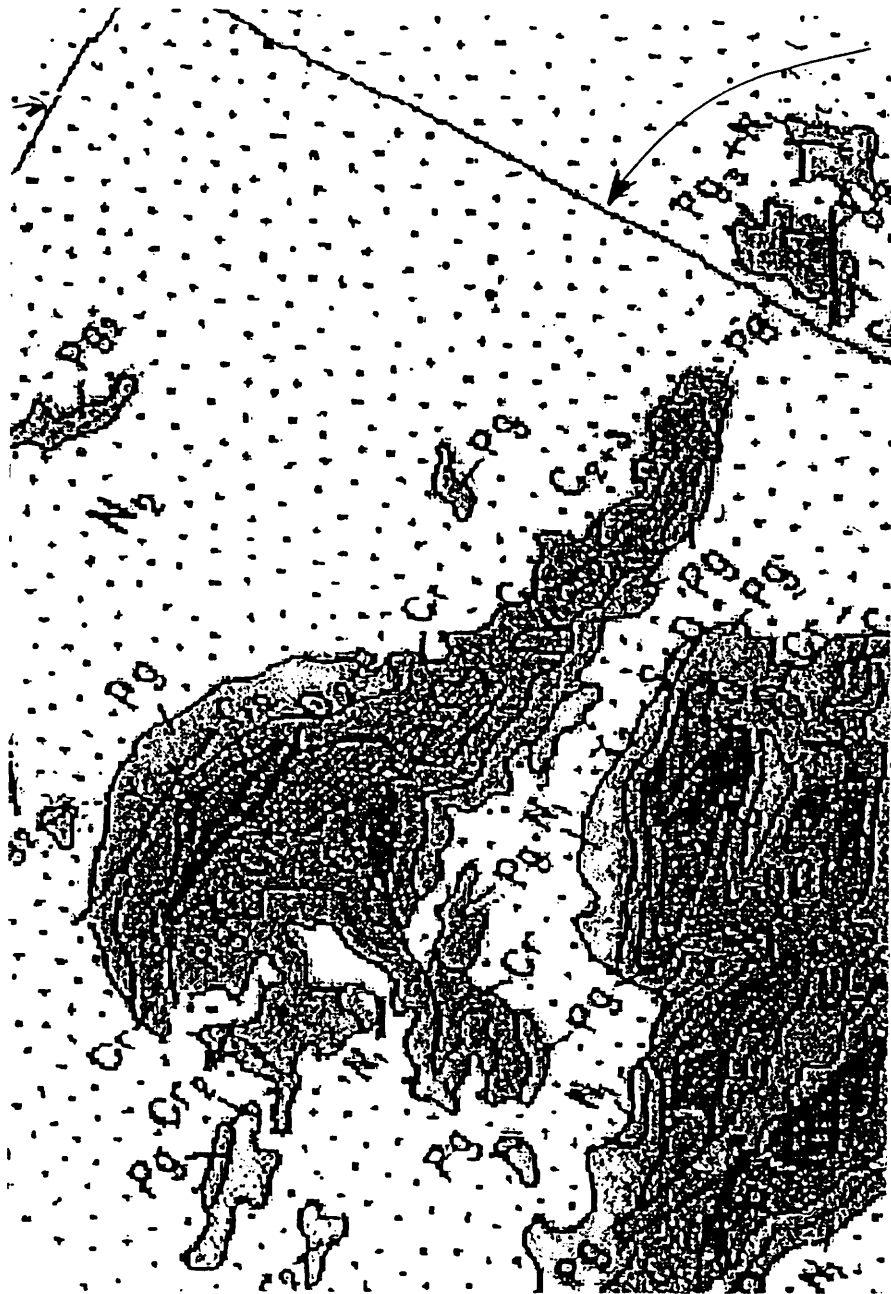


FIG. 1 – A fragment of a 1:7,500,000 scale geologic map of relatively poor quality, i.e. pale colors, variable line width, tiny polygons. Gray-scale scanned image of this fragment has been converted to grid by IMAGEGRID command and used as input for GRIDVECTOR program. Arrows point to a latitude/longitude lines that are considered to be noise. Other noise features are text labels and background dot patterns.



FIG. 2 – An intermediate average slope grid, from map fragment in figure 1, where the white pixels are maximums. Latitude/longitude grid lines and many of the text labels have disappeared.

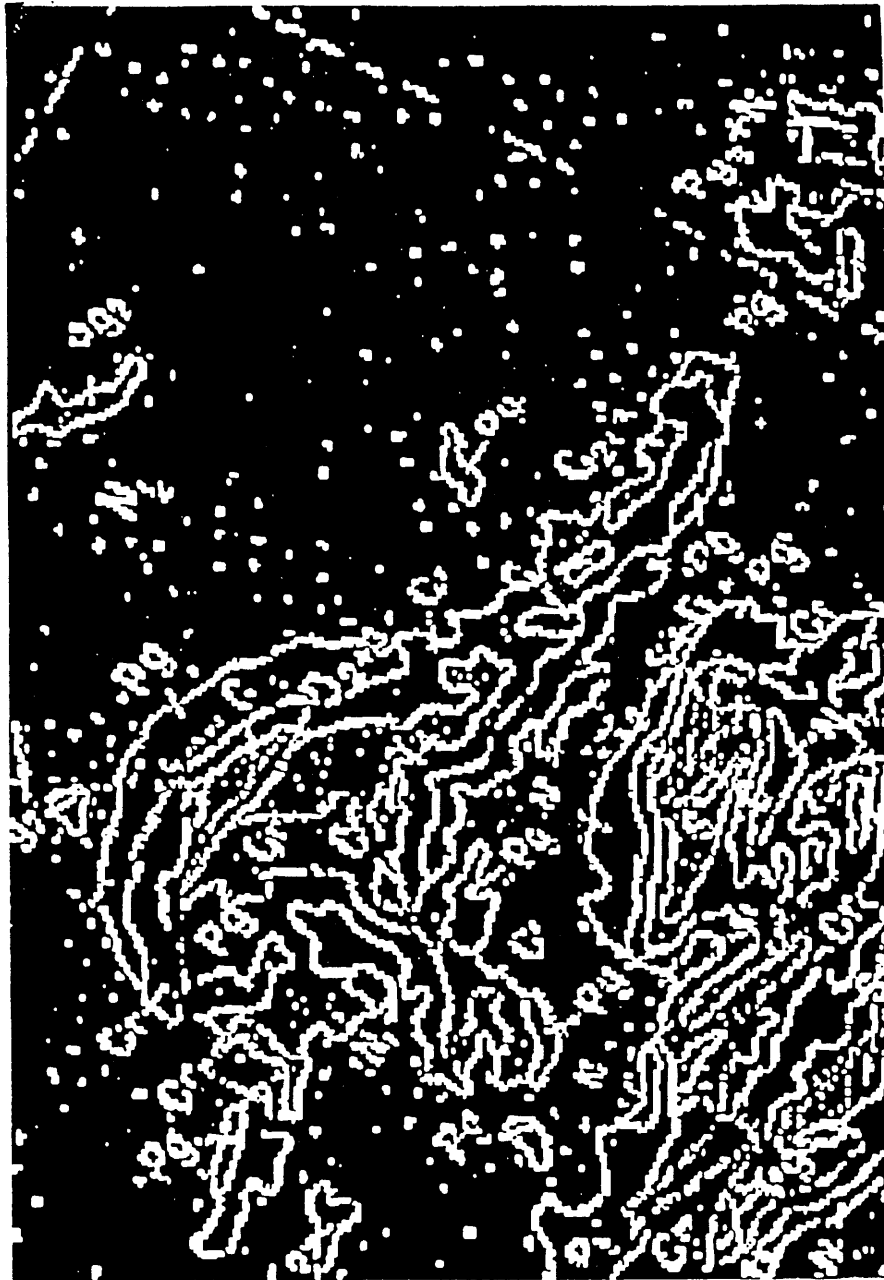


FIG. 3 – Binary grid, from map fragment in figure 1, after completion of input grid processing which serves as the input for GRIDLINE. Noise features including latitude/longitude grid lines are converted to groups of speckles that are in turn removed by the GRIDDESPECKLE command of ARC/INFO.

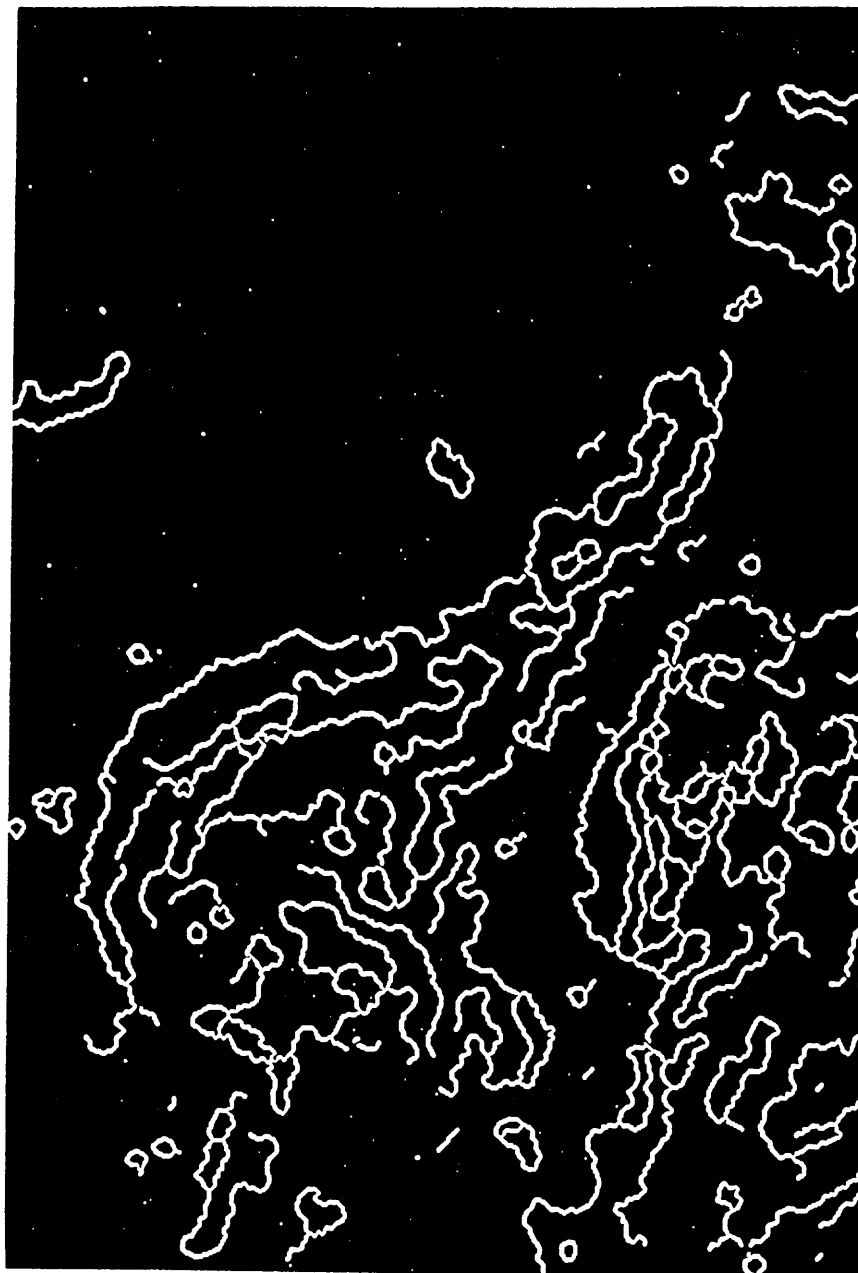


FIG. 4 – Final line and polygon ARC/INFO coverage after GRIDLINE. Coordinate lines and many of the text labels are completely filtered out. Some of the polygons look completed and are ready for geological attributing. The results in the lower-left corner however are of poor quality and further editing and correcting needs to be done before geological attributing can start.

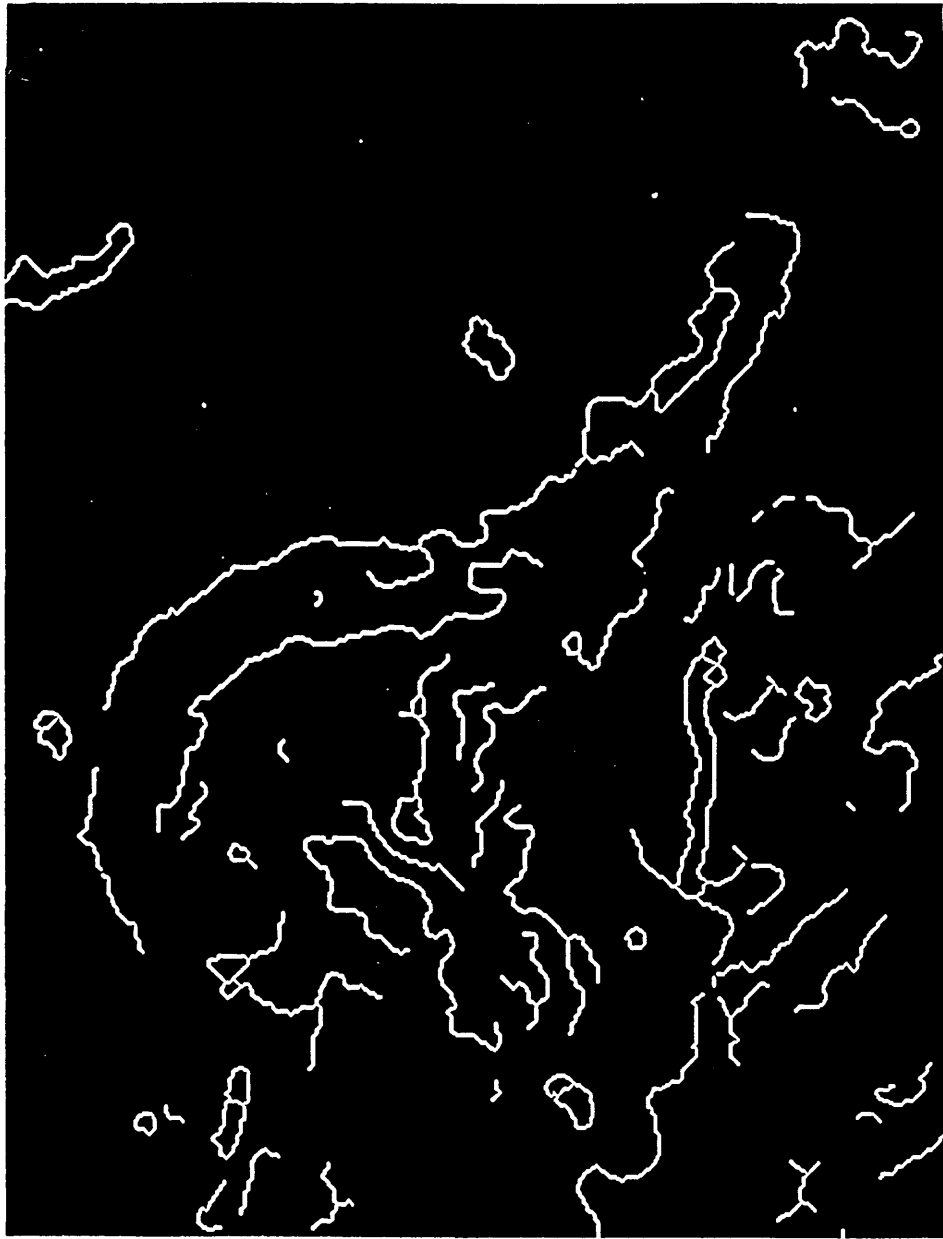


FIG. 5 – Another run on the same grid, using different parameters, a higher threshold for dark lines and average slope. There are fewer false lines in this coverage, but at the same time more digitization is needed.



FIG. 6 – A fragment of a map at 1:7,500,000 scale showing sea depth contours to be extracted. Areas between the contours are filled in by different patterns.

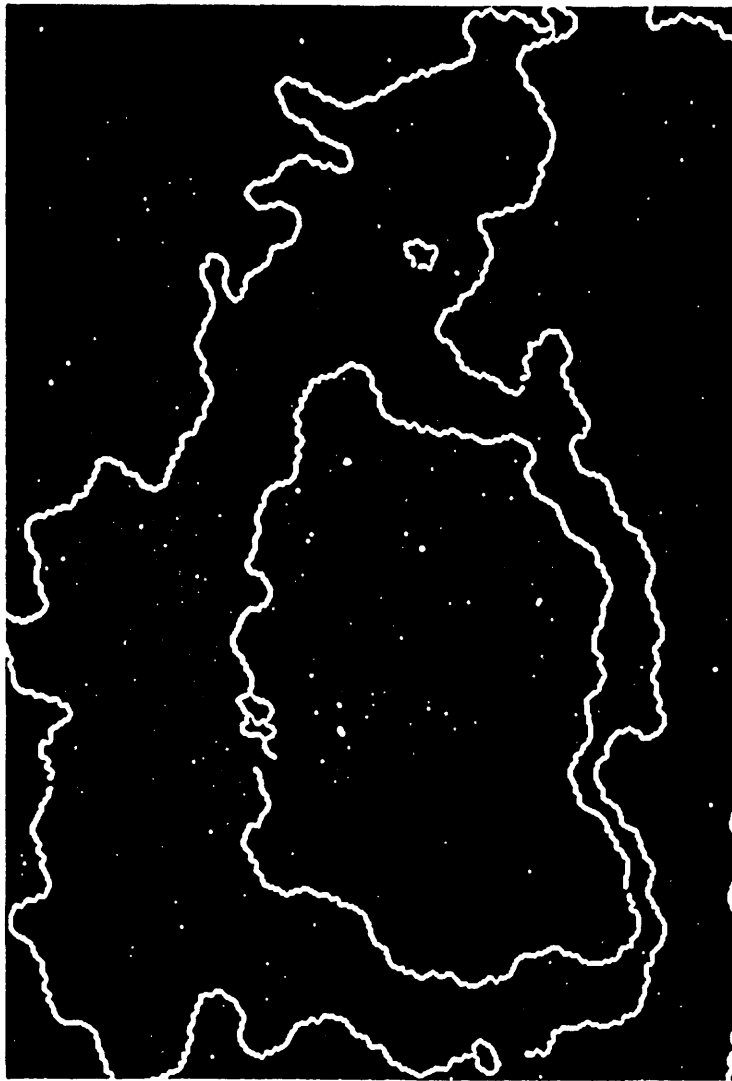


FIG. 7 – Final line ARC/INFO coverage. Minor corrections, like filling in gaps and removing false lines, will still need to be done.

```

/*****
/*
/*          GRIDVECT
/*          Version 1.0
/*          PROGRAMMER: FELIKS M. PERSITS
/*          January 20, 1997
/* THIS AML PROGRAM IS COMPATIBLE WITH ARC/INFO, ver. 7.0.4 RUNNING ON
/* SUN 1000 WITH SOLARIS 2.5 OPERATING SYSTEM. IT IS DESIGNED TO EXTRACT
/* LINEAR FEATURES (VECTORS) FROM A GRAY-SCALE GRID. THAT GRID IS A RESULT
/* OF THE CONVERSION OF A GEOREFERENCED, SCANNED IMAGE OF A PAPER GEOLOGICAL
/* MAP, USING ARC/INFO'S REGISTER AND RECTIFY COMMANDS WITH POINT COVERAGE
/* OF LAT/LONG CROSSES FROM THE ORIGINAL PAPER MAP AS A REFERENCE.
/* PROGRAM MAY WORK FOR OTHER SIMILAR APPLICATIONS.
/* DESIRED LINES ON THE MAP ARE BOUNDARIES SEPARATING VARIOUS GEOLOGICAL
/* FEATURES, DISPLAYED ON THE IMAGE BY THIN DARK STRIPS OR THE AVERAGE GRAY
/* LEVEL OR EVEN PATTERN. FEATURES TREATED AS NOISE ARE LAT/LONG GRID LINES,
/* RIVERS, TEXT ANNOTATIONS, ETC. AFTER EXTRACTION OF LINES, THE PROGRAM
/* CREATES A POLYGON COVERAGE WITH TWO ADDITIONAL ITEMS:
/* GLG - CHARACTER ITEM TO DENOTE GEOLOGICAL NAME OF POLYGON, INDICATING ITS
/* TYPE.
/* G - INTERGER ITEM TO KEEP TRACK OF POLYGONS POPULATED WITH GLG OR NOT.
/* THE PROGRAM IS BASED ON STANDARD ARC/INFO GRID, ARCSCAN AND ARCEDIT COMMANDS
/* BECAUSE THE PROGRAM USES THE ARC/INFO GRIDLINE COMMAND TO CREATE A LINE
/* COVERAGE FROM A BINARY GRID PREPARED BY THE PROGRAM, FINAL RESULTS ARE
/* VERY DEPENDENT ON HOW GRIDLINE WORKS.
/* OVERALL RESULTS ARE A FUNCTION OF THE PAPER MAP QUALITY, THE PARAMETERS
/* SELECTED, AND IMAGE RESOLUTION. BECAUSE OF ARCGRID'S MAXIMUM FILTER SIZE OF
/* 7 PIXELS, OPTIMAL SCANNER RESOLUTION IS SET TO 200 DOTS/INCH.
/* THERE ARE TWO MAIN CHOICES WHEN YOU DECIDE UPON THRESHOLDS FOR THE DARK LINE
/* VALUE AND THE AVERAGE SLOPE LEVELS:
/* 1. SELECT LOWER THRESHOLDS. IT RESULTS IN MORE LINES BEING EXTRACTED, BUT
/*    REQUIRES THE SUBSEQUENT REMOVAL OF MANY FALSE LINE SEGMENTS.
/* 2. SELECT HIGHER THRESHOLDS. IT RESULTS IN LESS LINES EXTRACTED, BUT MORE
/*    DIGITIZATION IS REQUIRED.
/* -----
/* !!! YOU HAVE TO HAVE ENOUGH DISK SPACE (AT LEAST THREE TIMES MORE THAN
/* THE INPUT GRID SIZE) BECAUSE THE PROGRAM CREATES SEVERAL TEMPORARY GRIDS.
/* -----
/* INPUT FILE - ARC/INFO GRAY-SCALE GRID.
/* OUTPUT FILE - ARC/INFO POLYGON COVERAGE "NEWLINES"
/* -----
/* VARIABLE LIST:
/*      inp      - name of input grid
/*      scle     - scale of original paper map
/*      smofilt  - smoothing filter size (3, 5, 7)
/*      enhfilt  - boundary enhancing filter size (3, 5, 7)
/*      minloc   - minimum value to select dark lines
/*      locslope1 - minimum value to select average slope
/*      s_mode   - mode of line extraction:
/*          1     - extract all possible lines
/*          2     - extract only dark lines
/*          3     - extract only gray-scale level changes
/*      wed      - weed tolerance for GRIDLINE and ARCEDIT
/*      thickn   - thickness for GRIDLINE
/*      grn      - grain tolerance for SPLINE
/*      edtdis   - editdistance for ARCEDIT
/*      ndssnap  - editsnap for ARCEDIT
/*      snp      - editsnap for ARCEDIT
/*      asnp     - arcsnap for ARCEDIT
/* -----
/* COMPLEMENTARY FILES IN WORKSPACE:

```

```

/*          filtset.menu  - menu file to select filters size and map scale
/*          minset.menu   - menu file to select dark line level
/*          slopeset.menu - menu file to select maximum average slope value
/*          s_mode.menu   - menu file to select mode for line extraction
/*-----
/*****
/*-----
&terminal 9999
display 9999 3
/*-----
/*-- Remove intermediate XX-files which might remain after previous runs.
/*-----
&type trying to delete /* XX-files
\rm xx*
\rm -r xx*
/*-----
/*--Check existence of input grid and menu files
/*-----
grid
clear
&sv inp := [getgrid '*' 'Enter input grid']
&if [exist %inp% -grid] and [exist filtset.menu -file] and [exist minset.txt ~
-file] and [exist slopeset.menu -file] and [exist s_mode -file] &then
&if [exist maj -grid] &then kill maj all
&if [exist smo -grid] &then kill smo all
&if [exist inpg -grid] &then kill inpg all
&if [exist inpgc -grid] &then kill inpgc all
&if [exist enh -grid] &then kill enh
&if [exist locdif1 -grid] &then kill locdif1 all
&if [exist locsl -grid] &then kill locsl all
&if [exist locslope -grid] &then kill locslope all
&if [exist locslope1 -grid] &then kill locslope1 all
&if [exist locslope2 -grid] &then kill locslope2 all
&if [exist filtloc -grid] &then kill filtloc all
&if [exist dif -grid] &then kill dif all
&if [exist loc -grid] &then kill loc all
&if [exist newl -cover] &then kill newl all
&if [exist newln -cover] &then kill newln all
&if [exist newlines -cover] &then kill newlines all
&do
mapextent %inp%
gridpaint %inp% # # # GRAY
/*-----
/*--Create rectangular grid by clipping input grid
/*-----
&type Select clipping polygon for all further processing
gridclip %inp% inpgc *
clear
mapextent inpgc
gridpaint inpgc # # # GRAY
/*-----
/* Thread to select map scale and filter size
/*-----
&thread &create filter_select &modal &menu filtset &pos &right &display
/*-----
/* do majority filter on input grid to create smoother mean
/*-----
inpg = majorityfilter(inpgc,four,majority)
&type draw maj - smoothing
shadeset colorrange

```

```

        gridshades inpg # linear
/*-----
/* do max smoothing filter twice
/*-----
        &type Start grid processing
        gridedit drawing off
        gridedit edit inpg
        &type First smoothing
        gridedit filtersize %smofilt%
        gridedit selectall
        gridedit smooth max
        &type the first smoothing done (smol)
        gridedit smooth max
        gridedit save smo
        gridedit drawing off
/*-----
/* do 3 times edge enhancement for input
/*-----
        &type start enhancement
        gridedit drawing off
        gridedit edit inpg
        gridedit filtersize %enhfilt%
        gridedit selectall
        gridedit enhance sharpen
        &type the first enh done
/* ! second edge enhancement
        gridedit enhance sharpen
        &type the second enh done
/* ! the third edge enhancement
        gridedit enhance sharpen
        &type the third enh done
        gridedit save enh
        kill inpgc all
        gridedit drawing on
        &type draw input enhanced
        shadeset colorrange
        gridshades enh # linear
/*-----
/*-- Do difference = majority filter on input - smoothed input
/* to select local min and max
/*-----
        &type Difference calculation
        gridedit drawing off
        dif = inpg - smo
        gridedit drawing on
        &type draw dif
        shadeset colorrange
        gridshades dif # linear
/*-----
/*--Interactive setting %minloc% value
/*-----
        &type SELECT minimum VALUE representing lines you need
        cellvalue dif *
        &thread &create line_val &modal &menu minset &pos &right &display
/*-----
/*--Do average slope on smoothed grid
/*-----
/* locdif1 = con(dif.value < -70,1,0)
/*-----
        kill inpg all

```



```

&type calc slope
locslope1 = focalmean(slope(smo,degree),rectangle,%smofilt%,%smofilt%)
kill smo all
&type slope smoothing
gridedit drawing on
shadeset colorrange
gridshades locslope1 # linear
/*-----
/* Interactive setting %locslope1% value
/*-----
    &type SELECT minimum VALUE representing lines on this grid
    cellvalue locslope1 *
    &thread &create slope_set &modal &menu slopeset &pos &right &display
    clear
/*-----
/*--Set of conditions to create binary grid "filtloc" as a input
/* for GRIDLINE
/*-----
    &thread &create s_mode &modal &menu s_mode &pos &right &display
    clear
    &select %s_mode%
    &when 1
        &do
            &type Mode is all
            if (dif <= %minloc% and enh <= 0.1 and ~
                (abs(locslope1) >= %maxslope%)) then
                filtloc = 1
            &end
        &when 3
            &do
                &type Mode is level
                if (abs(locslope1) >= %maxslope%) then
                    filtloc = 1
                &end
            &when 2
                &do
                    &type Mode is dark
                    if (dif <= %minloc% and enh <= 0.1) then
                        filtloc = 1
                    &end
                &end
            &end
    &end
/*-----
    kill enh all
    kill dif all
    kill locslope1 all
    &type draw filtloc
    gridpaint filtloc
    &end
/* &else
/* &do
/* &type %inp% is not grid or you missed filtset, minset, ~
/* slopeset, s_mode menu files
/* &return
/* &end
/*-----
/*----- Exit from GRID
/*-----
q
/*-----
/*-----start GRIDLINE-----

```

```

/*-----
&type speckle removing
griddespeckle filtloc loc 7 7
kill filtloc all
display colormap default
&type start GRIDLINE
&sv wed = %scle% / 15000
&sv grn = %scle% / 15000
&sv edtdis = %scle% / 1500
&sv ndsnap = %scle% / 7500
&sv snp = %scle% / 7500
&sv asnp = %scle% / 10000
&sv thickn = %scle% / 250
gridline loc newlines data thin nofilter round # %thickn% 0.0 %wed%
kill loc all
/*-----
/*----- coverage processing-----
/*-----
display 9999 3
ae
ec newlines
ef lines
editdistance %edtdis%
grain %grn%
weedtolerance %wed%
nodesnap closest %ndsnap%
arcsnap on %asnp%
snapping closest %snp%
drawenv arc
symbolitem newlines arc 3
select all
spline method default
/* spline method mcconalogue
spline
select all
move parallel 0
intersect all
image %inp%
draw
save
q
/*-----
/*-----Create outer boundary for coverage
/*-----
&type draw outer frame
tables
sel newlines.bnd
&sv xmin = [show record 1 xmin]
&sv ymin = [show record 1 ymin]
&sv xmax = [show record 1 xmax]
&sv ymax = [show record 1 ymax]
q
ae
ec newlines
mape newlines
ef line
drawenv arc
symbolitem newlines arc 3
image %inp%
draw

```

```

&pushpoint 2 %xmin% %ymin%
&pushpoint 2 %xmin% %ymax%
&pushpoint 2 %xmin% %ymax%
&pushpoint 2 %xmax% %ymax%
&pushpoint 2 %xmax% %ymax%
&pushpoint 2 %xmax% %ymin%
&pushpoint 2 %xmax% %ymin%
&pushpoint 2 %xmin% %ymin%
&pushpoint 9 0 0
add
save
q
/*-----
/*--Clean, build coverage and eliminate small polygons
/* caused by GRIDLINE
/*-----
&type cleaning newlines
clean newlines newln 0.0 %wed%
kill newlines all
&type build newlines
build newln
eliminate newln newlines # poly # area
~res area < 15000000
~
n
n
kill newln all
/*-----
/*--Create additional items in polygon FAT and set default value
/* for item g = 3
/*-----
createlabels newlines
build newlines
additem newlines.pat newlines.pat glg 5 5 C
additem newlines.pat newlines.pat g 2 2 I
ae
make newlines
ec newlines
ef label
select all
calc g = 3
/*-----
/*--Draw final coverage and save result
/*-----
drawenv arc label
symbolitem newlines arc 3
symbolitem newlines label g
image %inp%
draw
&pause 'Look at the result and hit ENTER to save'
save
q
&return

```

```

/*-----FILTSET.MENU-----
7
    Specify minimum value, representing line
Min_value_for_line: %minloc
    %apply
%minloc slider minloc 60 step 2 init -10 integer -150 0
%formopt setvariables immediate
%forminit &s minloc = -10
%apply button keep Apply &return
7
    Specify filter size and grid scale
Smoothing_filter: %smofilt
Boundary_filter: %enhfilt
Grid scale: %scle
    %apply
/* %smofilt choice smofilt single init 5 help 'Smooth filter =' 3 5 7
/* %enhfilt choice enhfilt single init 5 help 'Boundary filter =' 3 5 7
%scle input scle 10 init 7500000 help 'Scale denominator' integer
%smofilt input smofilt 10 init 5 help 'Smooth filter (3,5,7)' integer
%enhfilt input enhfilt 10 init 5 help 'Boundary filter (3,5,7)' integer
%formopt setvariables immediate
%forminit &s smofilt = 5
%forminit &s enhfilt = 5
%apply button keep Apply &return
/*-----MINSET.MENU-----
7
    Specify minimum value, representing line
Min_value_for_line: %minloc
    %apply
%minloc slider minloc 60 step 2 init -10 integer -150 0
%formopt setvariables immediate
%forminit &s minloc = -10
%apply button keep Apply &return
/*-----SLOPESET.MENU-----
7
    Specify maximum slope for boundary
Max_value_for_slope: %maxslope
    %apply
%maxslope slider maxslope 40 step 0.02 init 0.1 real 0 2.0
%formopt setvariables immediate
%forminit &s maxslope = 0.08
%apply button keep Apply &return
/*-----S_MODE.MENU-----
7
    Select source of lines:3-gray level jump, 2-dark outline or 1-all sources
Mode: %s_mode
/* Boundary_filter: %enhfilt
/* Grid scale: %scle
    %apply
%s_mode choice s_mode single init all help 'What to trace=' 1 2 3
/* %enhfilt choice enhfilt single init 5 help 'Boundary filter =' 3 5 7
/* %scle input scle 10 init 7500000 help 'Scale denominator' integer
/* %smofilt input smofilt 10 init 5 help 'Smooth filter (3,5,7)' integer
/* %enhfilt input enhfilt 10 init 5 help 'Boundary filter (3,5,7)' integer
%formopt setvariables immediate
%forminit &s s_mode = 1
/* %forminit &s enhfilt = 5
%apply button keep Apply &return

```