AD-A095 986 ANDRULIS RESEARCH CORP BETHESDA MD F/G 8/2
AUTOMATIC CARTOGRAPHIC FEATURE IDENTIFICATION.(U)
JAN 81 T J LAWSON F30602-79-F-0215
UNCLASSIFIED RADC-TR-80-387 NL

1 OF 1
AD A
095986

END
DATE
FILMED
4-81
DTIC

① LEVEL II

AD A095986

RADC-TR-80-387
Final Technical Report
January 1981

AUTOMATIC CARTOGRAPHIC FEATURE IDENTIFICATION

Andrulis Research Corporation

Thomas J. Lawson

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DTIC
ELECTE
MAR 4 1981
S D
B

DDC FILE COPY

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

81 3 3 056

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-80-387 has been reviewed and is approved for publication.

APPROVED: J R Baumann

JOHN R. BAUMANN
Project Engineer

APPROVED: Owen R. Lawter

OWEN R. LAWTER, Colonel, USAF
Chief, Intelligence and Reconnaissance Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (IRRP) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return this copy. Retain or destroy.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: RADC-TR-80-387
2. GOVT ACCESSION NO.: AD-A095986
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): AUTOMATIC CARTOGRAPHIC FEATURE IDENTIFICATION
5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report Aug 79 — Sep 80
6. PERFORMING ORG. REPORT NUMBER: N/A
7. AUTHOR(s): Thomas J. Lawson
8. CONTRACT OR GRANT NUMBER(s): F30602-79-F-0215
9. PERFORMING ORGANIZATION NAME AND ADDRESS:

Andrulis Research Corporation
7315 Wisconsin Ave-650N
Bethesda MD 20014

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 63701B 32030307
11. CONTROLLING OFFICE NAME AND ADDRESS:

Rome Air Development Center (IRRP)
Griffiss AFB NY 13441

12. REPORT DATE: January 1981
13. NUMBER OF PAGES: 42
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report): Same
18. SUPPLEMENTARY NOTES: RADC Project Engineer: John R. Baumann (IRRP)
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Cartography
Digital Cartography
Contour Labelling
Feature Tagging

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The objective of the study efforts presented in this report was to identify and test automatic identification techniques whereby descriptor data is entered into the digital record for selected features and additional descriptor data for given sets of features is derived by computational techniques.

This study was divided into three phases. In the first phase, general

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE background and knowledge of the problems was gained, and potential solutions were identified. This was accomplished by site visits to DMA facilities, literature review, and discussions with experts in cartography and related areas. Section 2 of the report discusses this phase in greater detail. The second phase was concerned with experimental testing of techniques identified as having the greatest potential payoff. Section 3 of the report details the experiments conducted, their significance and recommended further testing in these areas. Section 4 contains more detail on suggested further research based upon the experimental results. The third and final phase was concerned with the development of a conceptual integrated system for the digital processing of cartographic data. Section 5 presents Andrulis' general concepts for the design of such a system.

TABLE OF CONTENTS

FOREWARD

LIST OF FIGURES

LIST OF TABLES

PREFACE

Andrulis Research Corporation (ANDRULIS), under the Sponsorship of the Rome Air Development Center (RADC) and the Defense Mapping Agency (DMA), has completed a study and analysis for the development of automatic cartographic feature identification. This effort was conducted under contract to the U.S. Air Force Systems Command RADC, contract number F30602-79-F-0215/subcontract SB3-4-7-8(a)79-C-769. This report presents an overview of the study, its results, and recommendations for further effort. It was prepared by Thomas J. Lawson, ANDRULIS Program Manager.

EVALUATION

High speed digitization of cartographic data has been hampered by a lack of correspondingly high speed methods for identifying feature information in the data files. This effort represents a promising beginning to the solution of this problem.

There is an immediate need by the Defense Mapping Agency for the types of feature tagging addressed under this contract. Future work should lead to successful production implementation of feature tagging systems.

John R Baumann

JOHN R. BAUMANN
Project Engineer

SECTION 1
EXECUTIVE SUMMARY

1.1 OBJECTIVE OF THE STUDY

The objective of the study efforts presented in this report was to identify and test automatic identification techniques whereby descriptor data is entered into the digital record for selected features and additional descriptor data for given sets of features is derived by computational techniques.

1.2 STUDY PHASES

This study was divided into three phases. In the first phase, general background and knowledge of the problems was gained, and potential solutions were identified. This was accomplished by site visits to DMA facilities, literature review, and discussions with experts in cartography and related areas. Section 2 of the report discusses this phase in greater detail. The second phase was concerned with experimental testing of techniques identified as having the greatest potential payoff. Section 3 of the report details the experiments conducted, their significance and recommended further testing in these areas. Section 4 contains more detail on suggested further research based upon the experimental results. The third and final phase was concerned with the development of a conceptual integrated system for the digital processing of cartographic data. Section 5 presents ANDRULIS' general concepts for the design of such a system.

1.3 RECOMMENDATIONS

1.3.1 INTEGRATED SYSTEM PLAN

ANDRULIS recommends that the DMA develop an integrated system plan for the automation of cartographic data processing. The plan is needed now, although the plan for implementation of such a system is in a much more distant timeframe. Without such a plan, a lack of compatibility between the stand-alone systems being acquired could seriously jeopardize DMA's ability to ever achieve a cost/effective, comprehensive system.

1.3.2 FURTHER RESEARCH EFFORTS

Based upon the conceptual system design presented in Section 5 of this report, the basic techniques developed and tested during this study should be developed into operational capabilities. Such techniques could greatly reduce the labor-intensity efforts currently required for digitizing cartographic data. Although complete automation might not be achievable in a foreseeable timeframe, a greater improvement is possible, and the labor-intensive portions remaining can be much more efficient by computer assistance. Section 4 of this report details the approach ANDRULIS recommends for the operational development of these techniques.

1.3.3 CHARACTER RECOGNITION

An idea that ANDRULIS feels is worthy of further study by DMA is the use of machine readable labels for manuscript preparation. Two different types of labels that have been operationally proven are the character type used on bank checks and the universal product codes used in computerized checkout for retail operations. A separate overlay for the labels would have to be scanned to establish registration and label locations. The character recognition hardware could then be guided into position to read and pass the data values into the system.

1.4 CONCLUSIONS

ANDRULIS is pleased with the results of their experimental efforts. No major dead-ends or false-starts were encountered. It was found that run-length encoded data was more efficient than vectorized data for the identification and location of the cartographic features tested. However, this could have been due to the type of system used and/or the types of features processed. It is felt that there is a very high potential payoff at a relatively low risk in continuing research and development in the area of automatic feature identification and tagging.

SECTION 2
PRELIMINARY RESEARCH

2.1 REVIEW OF CURRENT DATA CAPABILITIES

Members of the ANDRULIS project staff spent two days at DMAAC in St. Louis, MO. on 16 and 17 October 1979, and two days at DMAHTC in Washington D.C. on 18 and 19 October 1979.

2.1.1 HARDWARE ACQUISITION

It appears to ANDRULIS that one of the basic shortcomings in the DMA planning for upgrading their capabilities is a lack of an overall integrated system plan for the hardware acquisition. The sophisticated processing techniques needed to automate cartographic data processing require that upgrading occur in a sequential manner. However, it appears that the hardware for stand-alone type of applications is not being acquired as part of an integrated system plan allowing for direct compatability and interconnection of the separate functions. The lack of automated data transfer between the various hardware can slow down the data processing system and increase the labor requirements because of interstep data handling and storage requirements. Quality degradation can also occur as data is digitized on one system, converted to hardcopy for transfer to another system, then re-digitized on the second system. The integrated system plan should specify the input/output format in detail for all functions. This will insure that any proprietary restrictions are internal to the function and do not hamper its integration into the overall automated cartographic data processing scheme.

2.1.2 CARTOGRAPHIC/DATA PROCESSING EFFICIENCY

A basic philosophy that was apparent to ANDRULIS during its DMA site visits was the orientation of preparing all cartographic products for human readability. This pride in preparation of their products is obviously why DMA produces excellent cartographic material. However, when the materials being prepared are required to interface with automated equipment, the preparation philosophy needs to be oriented towards data processing efficiency rather then human readability. For example, if a DLMS manuscript is being prepared for data processing, it would be more efficient for processing purposes to use separate overlays and write the feature numbers for point, linear, and small areal features over top of the feature instead of off to the side with an arrow.

2.1.3 OTHER HARDWARE CONSIDERATIONS

It does not seem that software capabilities are being utilized to the fullest extent possible for assisting the human operator at the AGDS editing station. Automated editing techniques could greatly ease the workload at this position by detecting and resolving some of the editing problems and prompting the operator where problems are detected that cannot be resolved by the software. Human engineering considerations are important at this station. The operator while maintaining final approval control of all operations, should be relieved of the mundane, time-consuming tasks that degrade his or her efficiency.

2.2 STATE-OF-THE-ART SURVEY

ANDRULIS spent an extensive amount of effort reviewing the state-of-the-art in the general field of automated cartography. A staff member attended the Automated Cartography Symposium in November 1979. Many discussions were held between staff members and experts involved in related fields of endeavor. One of the project staff members was enrolled in a senior level cartography course at the University of Maryland. Extensive reviews of the literature were conducted, mainly using local area university libraries and the NASA/GODDARD Space Flight Center Library, as well as reports supplied by the Rome Air Development Center.

2.3 AREAS OF RESEARCH IDENTIFIED

Subsequent to the DMA site visits and general survey conducted, many areas for further study were identified.

2.3.1 EDIT ASSISTANCE

Based upon ANDRULIS' observations at the AGDS editor stations at DMA, noise removal was identified as a slow and labor-intensive problem. The observed problems in this area were in the form of breaks introduced into continuous lines and spurs and stray features introduced into the digitized data. It appeared that software could be developed for contour data that would automatically remove these types of noise in many cases, and prompt the editor for assistance in those situations where the solution was either not apparent or ambiguous to the software. Contour data was selected because of the basic rule that contours could neither begin nor end within the data set. The application of the digital Laplacian would emphasize the illegal end-points that occurred in such data.

2.3.2 GENERAL FEATURE IDENTIFICATION

It appears that Fourier and Hadamard transformation analysis could be utilized for the identification of features. An analysis of the spectra resulting from these transforms should permit the identification of characteristic components by which features can be identified. These patterns should correspond to line weight, length of dashes, size of dots, occurrence of tick marks, repetition of dots or dashes, and other visually unique characteristics. It was also recognized that a variety of pattern analysis techniques could be applied to the extraction and classification of the properties of features.

2.4 TESTS DEFINED FOR COMPUTER EXPERIMENTATION

After presenting these approaches to technical staff members, it was decided that in order to satisfy the most immediate needs of DMA, computer experimentation would be conducted in the area of identifying areal, lineal, and point features and associating labels with these general types of features. It was determined that contour and DLMS types of features would be the basis of this experimentation. The following section contains a detailed description of the experiments conducted by ANDRULIS.

SECTION 3
EXPERIMENTS CONDUCTED

3.1 CONCEPTUAL APPROACH

Four experiments were designed for testing and demonstrating automatic feature identification and label association. The first two experiments were conducted to establish the credibility of the concepts applied in the last two.

3.2 EXPERIMENT ONE, PROCESSING DMA SUPPLIED DATA

3.2.1 INPUT DATA

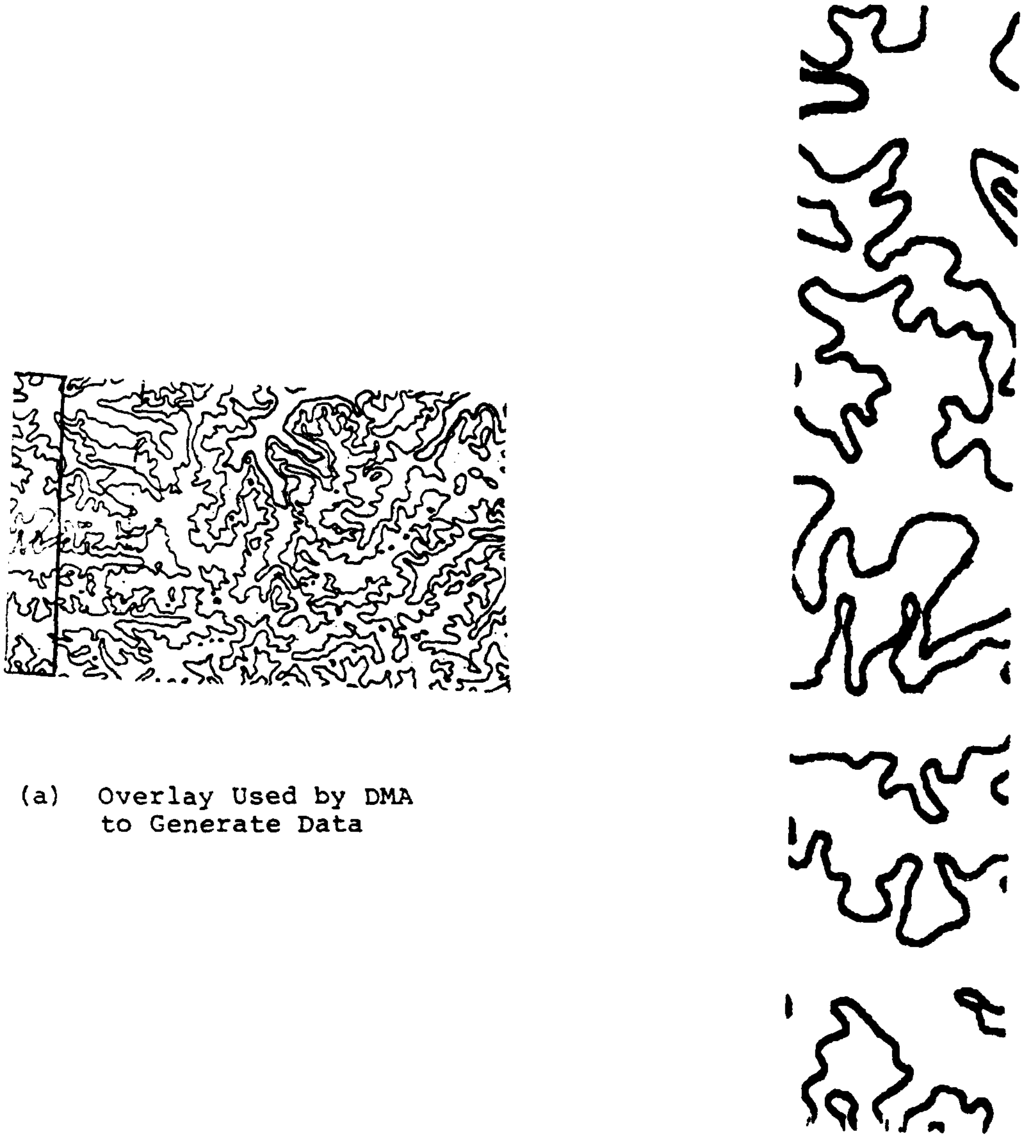
For this experiment, run-length encoded contour data generated by DMA on the AGDS System was used.

3.2.2 PROCESSING

The processing in this test reads the run-length encoded contour data and reproduces the plot on the graphic display. Only a small portion of the overlay used to generate the data is reproduced because of the slow I/O involved in the experiment. Figure 3.1 depicts this experiment. Figure 3.1 (a) is the plot used by DMAHTC to generate the contour data on the AGDS. Figure 3.1 (b) is the reproduction of the left portion of the overlay that was reproduced at ANDRULIS.

3.2.3 SIGNIFICANCE

Since the experiments concerned with the actual label and contour/feature association use idealized data, the significance of this experiment is its verification that ANDRULIS can read actual cartographic data generated by DMA.



(a) Overlay Used by DMA to Generate Data

(b) ANDRULIS Generated Plot

Figure 3.1: Processing DMA Supplied Data

3.2.4 RECOMMENDATIONS

The AGDS generated data for the contour overlay was supplied to ANDRULIS by DMAHTC in both raster (run-length encoded) and vectorized format. The plotting of the contours was done using the run-length encoded data, but both formats were read by ANDRULIS to verify that they could be used. Data generated from a label overlay should also be processed to verify that it can be utilized in the manner of the ANDRULIS created label data processed in Experiment Two.

3.3 EXPERIMENT TWO, LABEL PROCESSING

3.3.1 INPUT DATA

For this experiment, the input data was created by ANDRULIS. It consists of a vector representation of the individual characters that make up the labels.

3.3.2 PROCESSING

The basic processing flow is shown in Figure 3.2. After the vectorized character data is read into the program, the maximum and minimum x and y values and midpoint of each character are computed. The distances between the midpoints of the frames formed for each of the characters are then evaluated and associated to form label frames. A printed output is generated to report on the associations made to create the label boxes. For demonstration purposes, the original input data is displayed on the graphics CRT, see Figure 3.3. Next a plot of the individual character frames and the composite label frames are displayed, Figure 3.4. The next step in the process is for the location of each label to be passed to a character reading capability so that the value within the label frame can be scanned and passed back to the label processing capability. Since this effort is not concerned with character recognition, ANDRULIS utilized an old fashion, but highly efficient, character reading technique for test purposes. The position of each label is drawn on the screen, and the operator is asked to enter the value of the label. Figure 3.5 shows

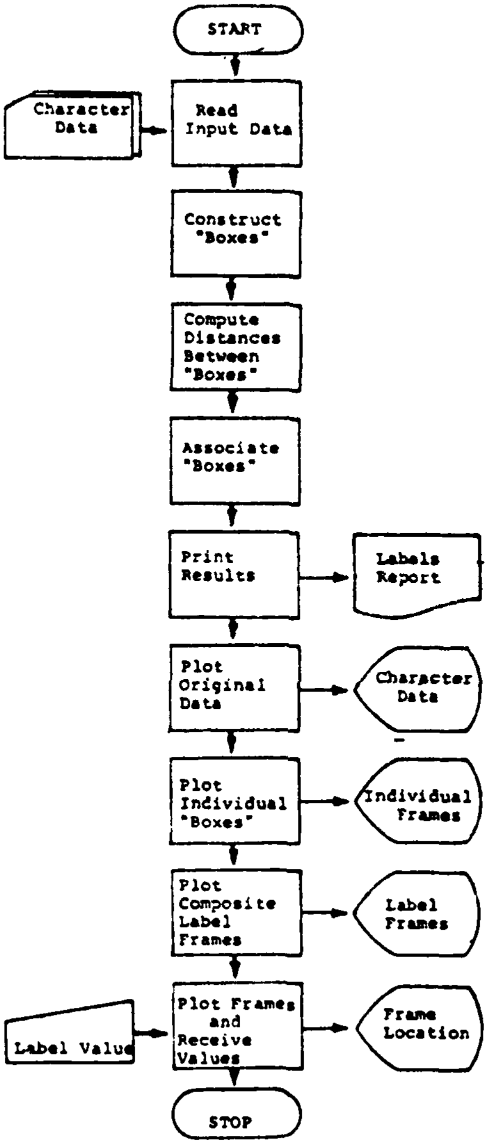


Figure 3.2: Label Data Processing

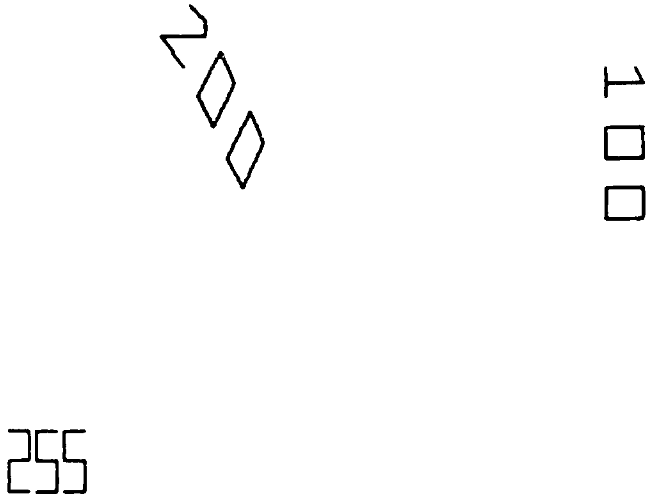
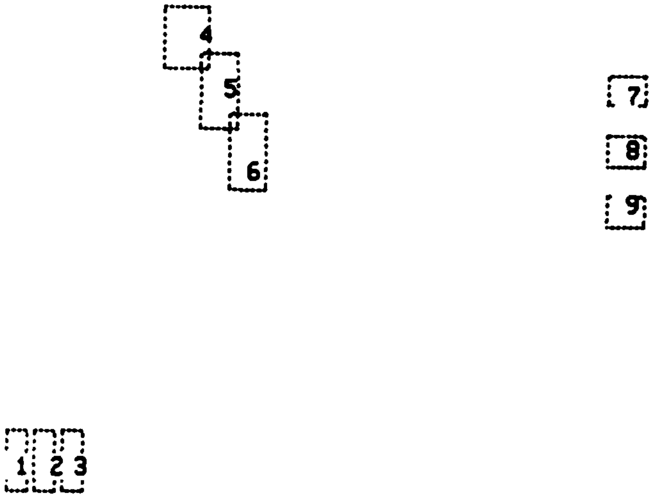
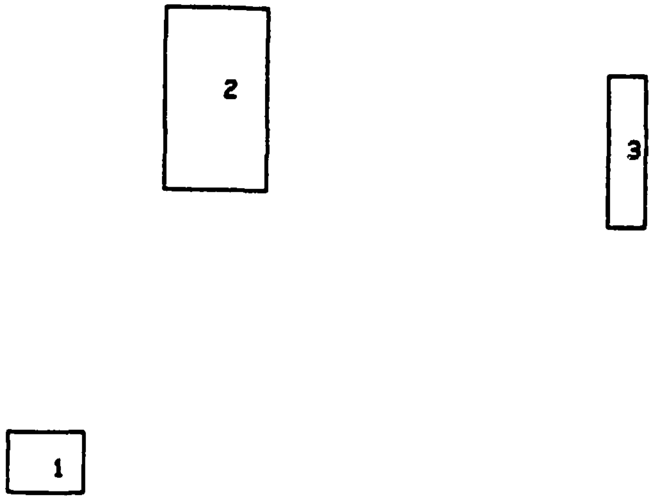


Figure 3.3: Character Data Processed



(a) Character Frames



(b) Label Frames

Figure 3.4: Charater and Label Frames

Figure 3.5: System Request for Label Value

the system identifying the location of a label and requesting that the value be entered. The value of each label is then entered on the final output report shown in Figure 3.6.

3.3.3 SIGNIFICANCE

This experiment demonstrates that it is possible to receive data created from a label overlay in a format that can currently be generated by DMA, process the data to locate the individual labels, send the location of each label to a character recognition capability, and receive the label value back from that capability. The method of creation of the labels is not a factor directly involved in the techniques used in this experiment, although it is a factor in the DMA's ability to create the digitized input data, and the character recognition system's ability to read the label once its position is specified.

3.3.4 RECOMMENDATIONS

The experiment conducted proves the concept. The ways that the manuscript can be created and converted to digitized data are untested. The labeling of contours that ANDRULIS has observed are carried out on a separate overlay and are handwritten onto the overlay. The label framing techniques should be tested using data created by the AGDS system from these overlays. This testing should also include the contour data with which the labels are associated. The testing will then include overlay registration alignment problems.

3.4 EXPERIMENT THREE, CONTOUR AND LABEL ASSOCIATION

3.4.1 INPUT DATA

For this experiment, the input data was created by ANDRULIS. The contour data file is in run-length encoded form. The label data consists of vectorized boxes framing the label locations. Since Experiment Two showed how the frames can be created from label data, those techniques are not repeated in this experiment.

```
COMPOSITE LABELS AND INPUT VALUES

LABEL  1   VALUE  255
LABEL  2   VALUE  200
LABEL  3   VALUE  100
      STOP


 STATUS OF CHARACTERS

RELATED CHARACTERS     1     2     3
RELATED CHARACTERS     4     5     6
RELATED CHARACTERS     7     8     9
```

Figure 3.6: Label Output Report

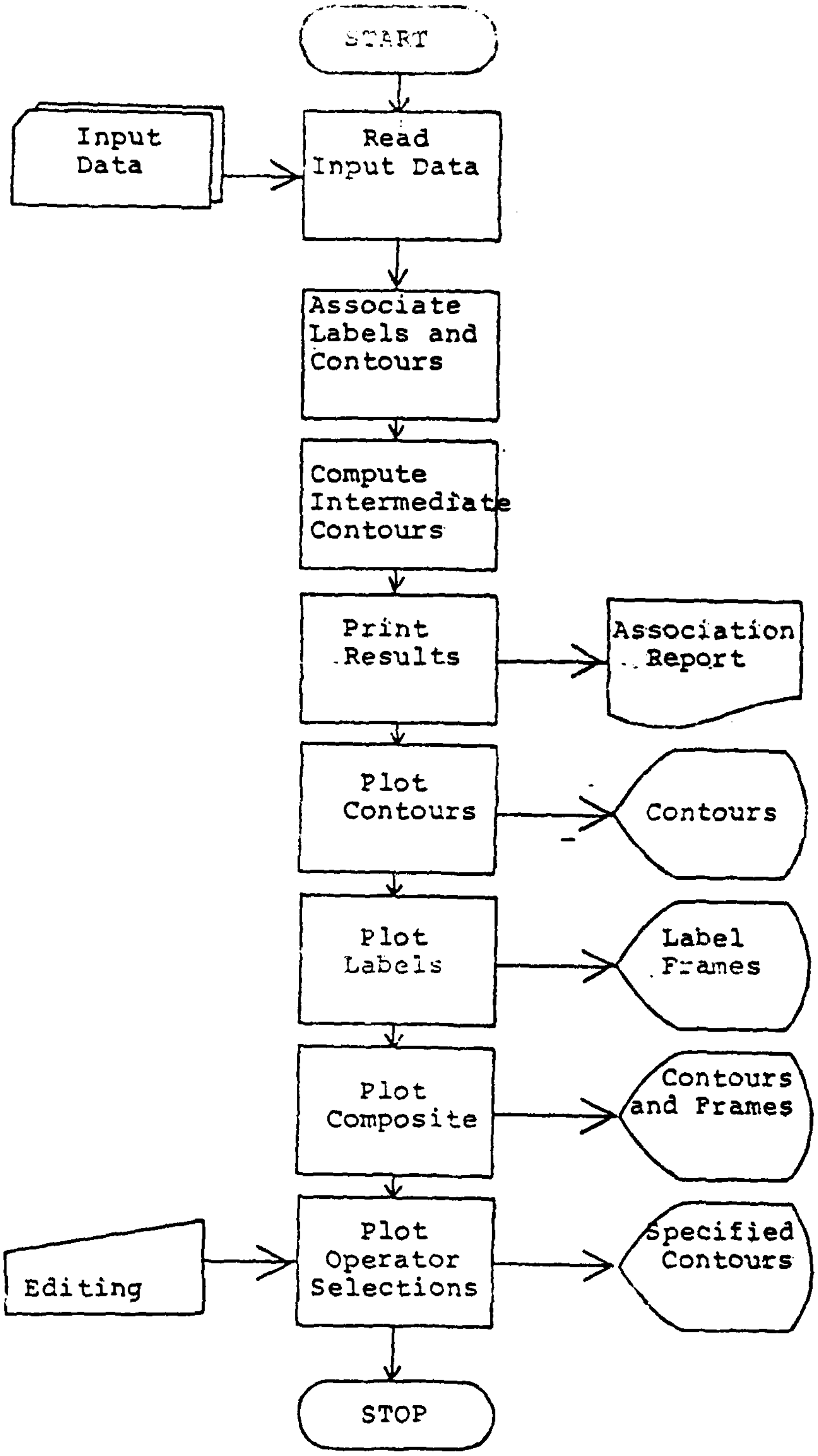


Figure 3.7: CONTOUR ASSOCIATION

```
CONTOUR  1 IS NOT ASSOCIATED WITH ANY LABELS
*** CONTOUR  2 HAS MULTIPLE LABELS
CONTOUR  2 ASSOCIATED WITH LABEL  1
CONTOUR  2 ASSOCIATED WITH LABEL  2
CONTOUR  3 IS NOT ASSOCIATED WITH ANY LABELS
CONTOUR  4 IS NOT ASSOCIATED WITH ANY LABELS
CONTOUR  5 IS NOT ASSOCIATED WITH ANY LABELS
CONTOUR  6 IS NOT ASSOCIATED WITH ANY LABELS
CONTOUR  7 ASSOCIATED WITH LABEL  5
CONTOUR  8 IS NOT ASSOCIATED WITH ANY LABELS
CONTOUR  9 ASSOCIATED WITH LABEL  4
CONTOUR 10 IS NOT ASSOCIATED WITH ANY LABELS
CONTOUR 11 IS NOT ASSOCIATED WITH ANY LABELS
CONTOUR 12 IS NOT ASSOCIATED WITH ANY LABELS



LABEL  1 ASSOCIATED WITH CONTOUR  2
LABEL  2 ASSOCIATED WITH CONTOUR  2
LABEL  3 IS NOT ASSOCIATED WITH ANY CONTOURS
LABEL  4 ASSOCIATED WITH CONTOUR  9
LABEL  5 ASSOCIATED WITH CONTOUR  7



   INDEX CONTOURS
INDEX  2      VALUE 200
INDEX  7      VALUE 100



   COMPLETE TABLE OF CONTOURS

          1 COMPUTED  VALUE 220
   INDEX  2 INPUT     VALUE 200
          3 COMPUTED  VALUE 180
          4 COMPUTED  VALUE 160
          5 COMPUTED  VALUE 140
          6 COMPUTED  VALUE 120
   INDEX  7 INPUT     VALUE 100
          8 COMPUTED  VALUE 240
          9 INPUT     VALUE 255
         10 COMPUTED  VALUE 240
         11 COMPUTED  VALUE 160
         12 COMPUTED  VALUE 140
```

Figure 3.8: Contours Association Report

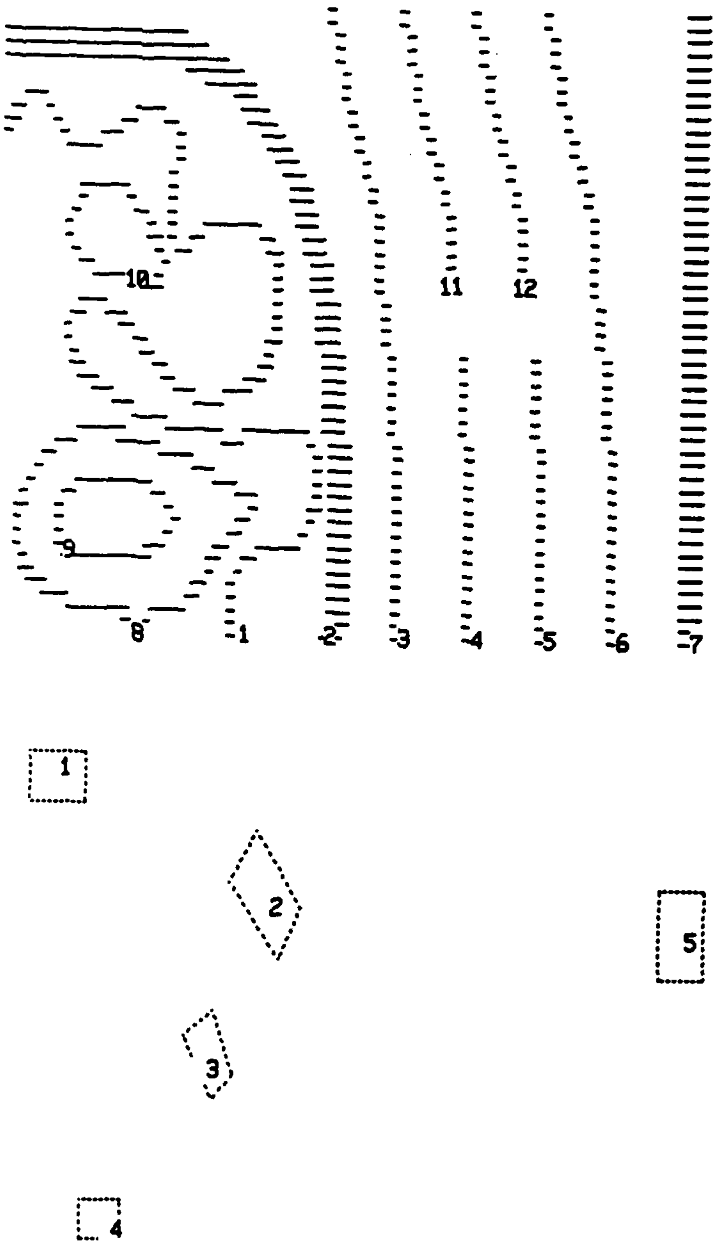


Figure 3.9: Contour and Label Input Display

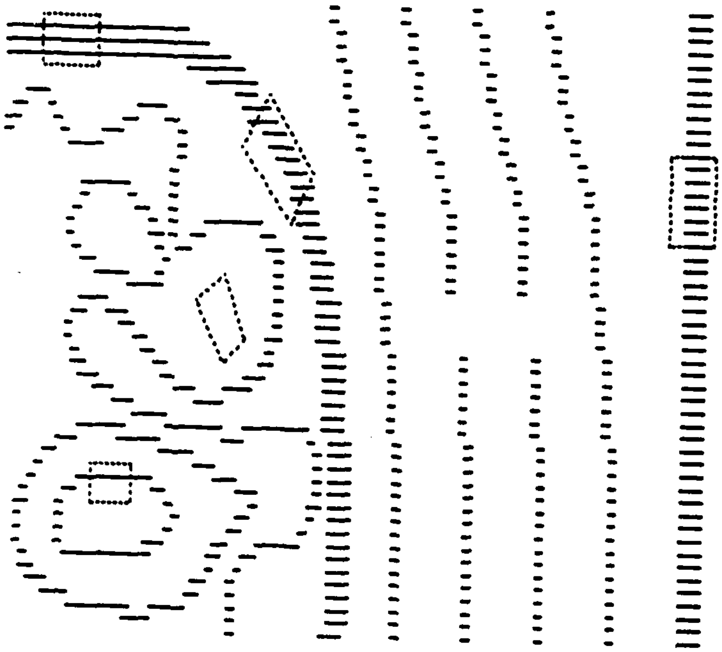


Figure 3.10: Labels Superimposed on Contours

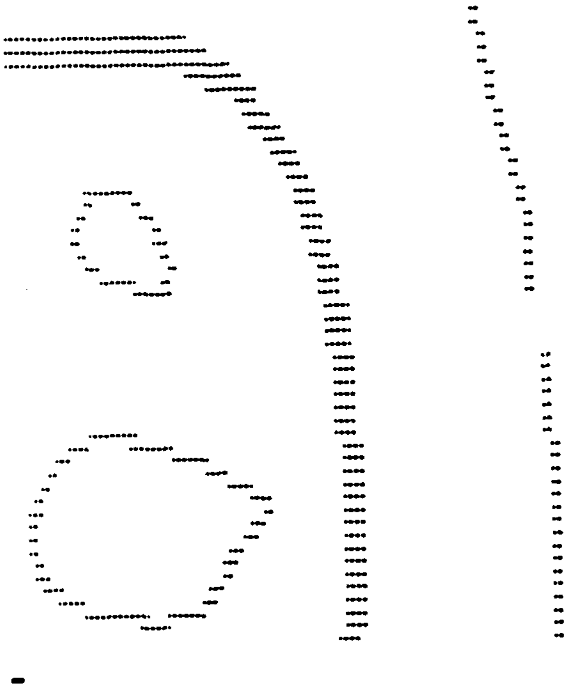


Figure 3.11: Operator Controlled Display of Selected Contours

3.4.2 PROCESSING

The basic processing flow is shown in Figure 3.7. After the input files are read into the program, the midpoint of each label frame is computed. For each frame, the shortest distance from the midpoint to each contour located within the input tolerance is computed. The frame is then associated with the contour that is closest and within the tolerance. Index contours are identified by noting the average thickness of each contour and comparing it to a hardcoded tolerance. A cross-reference of label values and frames are hardcoded into this experiment to represent the interaction between the label frame identification and obtaining of the label value from an exterior source that was shown in Experiment Two. For this experiment, the values of intermediate contours are interpolated or extrapolated based on the values of index contours and high/low point values given. The printed association report shown in Figure 3.8 is generated. Figure 3.9 shows the contour and label data with the temporary identification numbers assigned for report purposes. The association report indicates which contours have no labels, which have one label, and which have more than one label. For labels, it indicates which are not associated with any contours, which are uniquely associated with only one contour, and which are ambiguous in that they can be associated with more than one contour. Finally, there is a tabulation of the contours that identifies the type of contour, its value, and whether the value was an input or computed. The plotting section is independent of the computational section and is used for demonstration purposes. The operator has the capability of displaying selected contours of particular values. Figure 3.10 is a display of the labels superimposed on the contours. In Figure 3.11, the operator had requested that the 140,200, and 240 level contours be displayed.

3.4.3 SIGNIFICANCE

This experiment is the first one that demonstrated automatic feature identification; in this case, the features were run-length encoded contours. The process then went on to associate the labels with the features, computer label values for unlabeled features, and create a one-to-one tagging of labels to features.

3.4.4 RECOMMENDATIONS

The recommended further testing in paragraph 3.3.4 should be combined with further testing of the techniques demonstrated in this experiment. In the association report presented in Figure 3.8, the index contour identified as contour 2 had two labels associated with it. From the overlays observed by ANDRULIS, this is an acceptable labelling practice. However, software techniques should be developed that will obtain and compare the label values of multiple labeled contours to ensure that they are all equal. Also, the software should check that the labels associated with index contours are permissable values. The data supplied by DMA should be representative of typical densities to allow realistic tolerances for association to be developed. Automated techniques should also be developed and tested to perform contour editing to assist with the labor-intensive problem of editing broken contours and spurs.

3.5 EXPERIMENT FOUR, DLMS ASSOCIATION

3.5.1 INPUT DATA

For this experiment, the input data was created by ANDRULIS. The DLMS data is coded one feature at a time. Each feature contains an arbitrary ID created by the RELATE and LABPROC programs, the number of pairs, the minimum and maximum X and Y, and for each Y-value, the start X and stop X. The label data file contains an arbitrary ID created by the RELATE and LABPROC programs, the midpoint, and for each Y-value, the start X and stop X.

3.5.2 PROCESSING

The basic processing flow is shown in Figure 3.12. For demonstration purposes, the features are plotted with their arbitrary IDs (see Figure 3.13), then the label frames are plotted with their

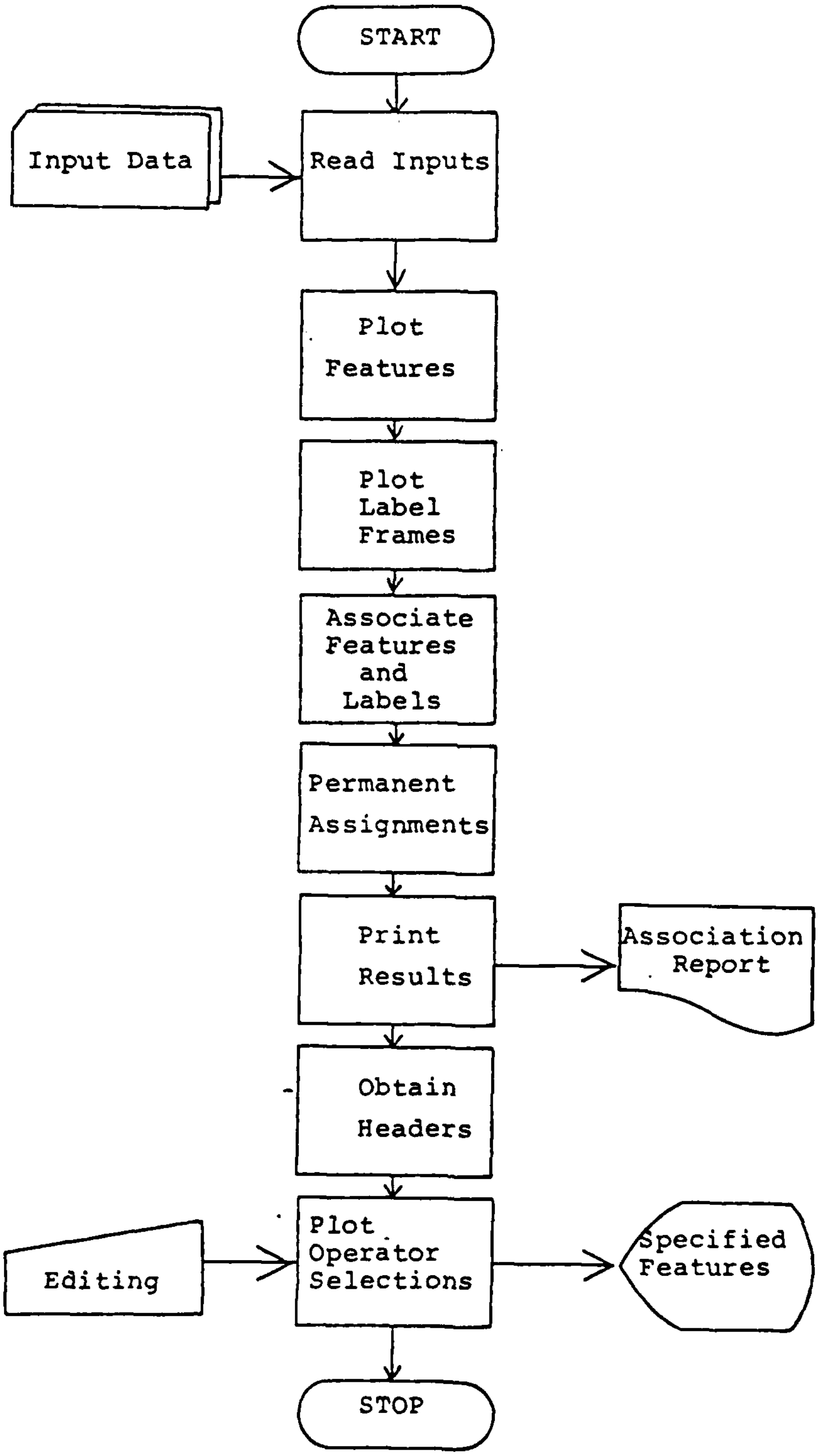


Figure 3.12: DLMS Association

Figure 3.13: Feature Data and Arbitrary IDs

arbitrary IDs (see Figure 3.14). Each label midpoint is checked to see if it falls within a feature or within a hardcoded input tolerance distance from a feature. Next each feature related to a label and each label related to a feature are stored. The feature array is then examined in an iterative fashion. All one-to-one associations of feature to label are eliminated and the label is eliminated from all other associations. This process continues until no more permanent associations can be made. The label array is then processed in the same manner. Next, the printed association report is created (see Figure 3.15). The printed report lists which feature is associated with each label using the arbitrary IDs, and also lists all ambiguous or non-associated labels. It then lists which label is associated with each feature along with ambiguous and non-associated features. Input data is used to simulate the identification of the actual label values, feature analysis code for DLMS processing. Once this value is obtained, the complete set of header data is available to the system from a source file created separately. The last portion of this experiment consists of a demonstration editing capability made available as a result of the association of the features with the header data. In the interactive editing mode, the operator can display features or labels by arbitrary ID. Based upon the header information, features can also be displayed by feature analysis code, type, surface material condition (including operator input groupings), and height by values or range of values. Figure 3.16 presents an operator called display of all areal (type 2) features, and Figure 3.17 was created when the operator request all features with an SMC between 3 and 7.

3.5.3 SIGNIFICANCE

The association techniques applied in this experiment extend the capability demonstrated in experiment three to include areal features, point features, and lineal features that start and end within the borders. The experiment goes one step further in its ability to associate a complete header with a feature.

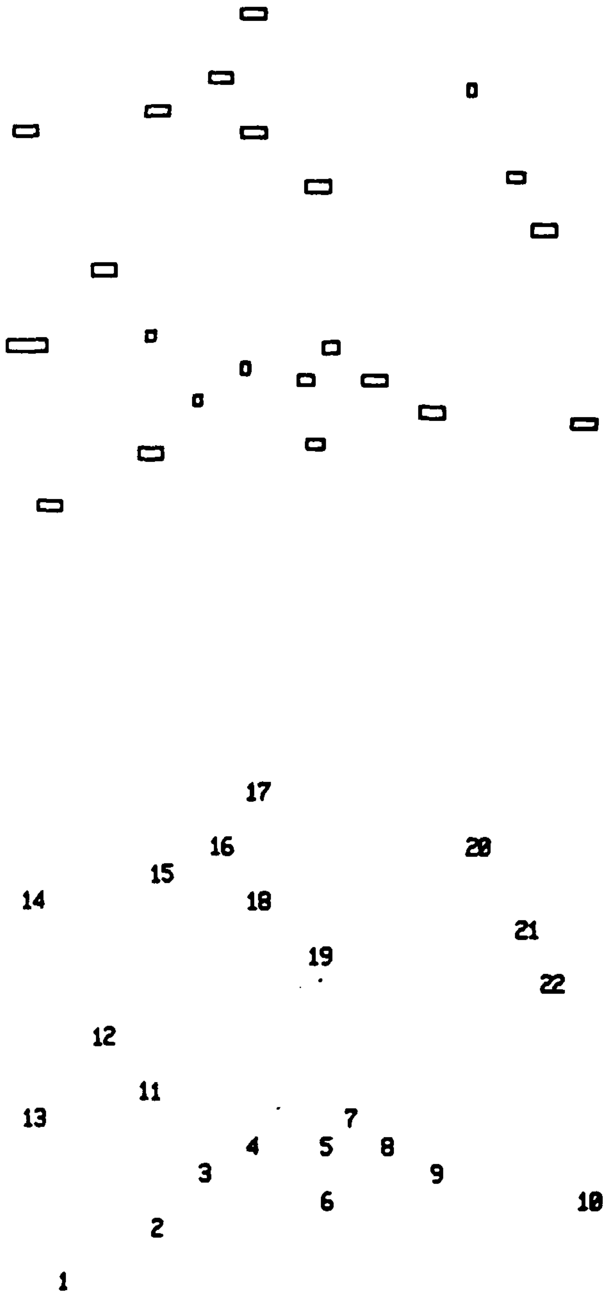


Figure 3.14: Label Frames and Arbitrary IDs

```
LABEL     1 IS ASSOCIATED WITH FEATURE    2

LABEL     2 IS ASSOCIATED WITH FEATURE    4

LABEL     3 IS ASSOCIATED WITH FEATURE    3

LABEL     4 IS ASSOCIATED WITH FEATURE    5

LABEL     5 IS ASSOCIATED WITH FEATURE    7

LABEL     6 IS ASSOCIATED WITH FEATURE    6

LABEL     7 IS AMBIGUOUS:

******

LABEL     7 IS ASSOCIATED WITH FEATURE    8

LABEL     7 IS ASSOCIATED WITH FEATURE    9

******

LABEL     8 IS ASSOCIATED WITH FEATURE   10

LABEL     9 IS ASSOCIATED WITH FEATURE   11

LABEL    10 IS ASSOCIATED WITH FEATURE   13

LABEL    11 IS NOT ASSOCIATED WITH ANY FEATURES

LABEL    12 IS ASSOCIATED WITH FEATURE   25

LABEL    13 IS ASSOCIATED WITH FEATURE    1

LABEL    14 IS ASSOCIATED WITH FEATURE   24

LABEL    15 IS ASSOCIATED WITH FEATURE   22

LABEL    16 IS ASSOCIATED WITH FEATURE   19

LABEL    17 IS ASSOCIATED WITH FEATURE   21

LABEL    18 IS ASSOCIATED WITH FEATURE   18

LABEL    19 IS ASSOCIATED WITH FEATURE   17

LABEL    20 IS NOT ASSOCIATED WITH ANY FEATURES

LABEL    21 IS ASSOCIATED WITH FEATURE   16

LABEL    22 IS ASSOCIATED WITH FEATURE   14

FEATURE     1 IS ASSOCIATED WITH LABEL   13
```

Figure 3.15: DLMS Association Report

```
FEATURE    2 IS ASSOCIATED WITH LABEL    1
FEATURE    3 IS ASSOCIATED WITH LABEL    3
FEATURE    4 IS ASSOCIATED WITH LABEL    2
FEATURE    5 IS ASSOCIATED WITH LABEL    4
FEATURE    6 IS ASSOCIATED WITH LABEL    6
FEATURE    7 IS ASSOCIATED WITH LABEL    5
FEATURE    8 IS ASSOCIATED WITH LABEL    7
FEATURE    9 IS ASSOCIATED WITH LABEL    7
FEATURE   10 IS ASSOCIATED WITH LABEL    8
FEATURE   11 IS ASSOCIATED WITH LABEL    9
FEATURE   12 IS NOT ASSOCIATED WITH ANY LABELS
FEATURE   13 IS ASSOCIATED WITH LABEL   10
FEATURE   14 IS ASSOCIATED WITH LABEL   22
FEATURE   15 IS NOT ASSOCIATED WITH ANY LABELS
FEATURE   16 IS ASSOCIATED WITH LABEL   21
FEATURE   17 IS ASSOCIATED WITH LABEL   19
FEATURE   18 IS ASSOCIATED WITH LABEL   18
FEATURE   19 IS ASSOCIATED WITH LABEL   16
FEATURE   20 IS NOT ASSOCIATED WITH ANY LABELS
FEATURE   21 IS ASSOCIATED WITH LABEL   17
FEATURE   22 IS ASSOCIATED WITH LABEL   15
FEATURE   23 IS NOT ASSOCIATED WITH ANY LABELS
FEATURE   24 IS ASSOCIATED WITH LABEL   14
FEATURE   25 IS ASSOCIATED WITH LABEL   12
```

Figure 3.15: DLMS Association Report (Continued)

Figure 3.16: Display of Operator Specified Areal Features

Figure 3.17: Display of all Features With SMC between 3 and 7

3.5.4 RECOMMENDATIONS

The basic techniques required for the identification and tagging of DLMS features have been demonstrated in this experiment. What is required next is a testing phase using data that has a representative density of features. This testing is needed to develop operationally oriented parametric and/or distributive functions for use in determining the most effective method of associating the labels with the features with minimum ambiguities and incorrect associations. Since the header data identifies the type of feature, a validation of all associations can be automated to insure that the type value listed in the header matches the feature characteristics, which can be established by an analysis of its vector representation.

3.6 GENERALIZED TECHNIQUES TESTED

For the association experiments, two generalized programs were created. They served as an interface between the raw input data and the association program.

3.6.1 RELATE PROGRAM

This program was designed to assign arbitrary IDs to either contour, feature, or label data. Its input data consists of one record for each Y-value. Each record contains the Y-value, the number of pairs of X-values, each start X and stop X pair. Arbitrary IDs are assigned to the first set of data read in and are written as the first record of an ID file. As subsequent sets of data are read, present pairs are compared to previous pairs. If the present pair is found to be related to a previous pair, it is given the same ID as the previous pair. If the same previous pair is found to be related to another present pair, the condition is flagged and written into a control file. This process continues until all Y-records are processed. One record is written to the ID file for each Y-value after that Y-value has been processed.

3.6.2 LABPROC PROGRAM

This program processes the ID file and control file prepared by the RELATE Program. The information is processed such that all the IDs that are flagged in the control file become associated. The individual IDs are compared against the associations. If an ID is found to be part of the association array, it is given a new ID. The new ID is the position at which the particular label, contour, or feature occurs. Finally, a new ID file is written with the arbitrary IDs in sequence, one for each label, contour, or feature.

3.7 GENERALIZED APPLICABILITY OF CARTOGRAPHIC DATA

The demonstrated ability to associate labels with line features, point features, and areal features that have been shown utilizing contour and DLMS data are applicable to other types of cartographic data. For example, with the data generated from the manuscript for a general area map, a line and associated label of 1002 can be tagged as a double track railroad, an area with a label of 4001 can be tagged as a swamp, and a point with a label of 0131 can be tagged as a mine. Hydrographic plots can be processed utilizing the same techniques as those applied to contour data.

SECTION 4
SUGGESTED FURTHER RESEARCH

4.1 AUTOMATED CONTOUR TAGGING

Given that a set of contours has been labeled automatically using label association and intermediate contour label calculation techniques, the problem of developing software to automatically connect broken contours and delete spurs becomes manageable. An example of how such a capability would work in a software environment is presented in this section.

Figure 4.1 presents a simple contour plot consisting of eight contours which have been labeled a to h, from top across and down, for discussion purposes. Table 4.1 shows the tagging that has been automatically performed using association techniques, given an overlay data set of tags for the two index contours, a and f. The label values for the six intermediate contours were calculated by the software. Contours c and g, and contours d and h are actually broken contours, and contour e contains a spur.

Figure 4.2 presents the portion of the contour plot that requires editing. Using a vector representation of the contours and the rule that all contours must begin and end on the edge of a plot, the five points that are of interest can be identified by software. They have been labeled from left to right for discussion purposes.

The next step in the process, starting with point 1, is to find the closest point, in this case, point 4. The label values for the two contours associated with the points, g and d, would be compared and found not to match. The process would then select the next closest point, 3, and again compare label values. Again the label values would not match and cause point 3 to be rejected. Point 2 would then be selected, a label match would be found, and a vector constructed to join contour c and contour g. As the process

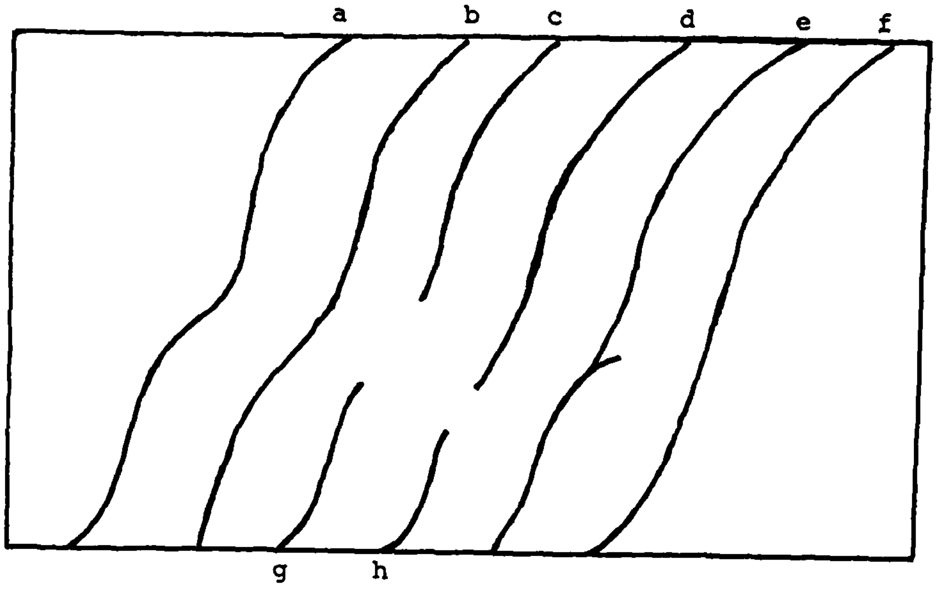


FIGURE 4.1: CONTOUR PLOT

Contour	Label Value	Source
a	100	associated
b	120	calculated
c	140	calculated
d	160	calculated
e	180	calculated
f	200	associated
g	140	calculated
h	160	calculated

TABLE 4.1: INTERNAL LABEL DEVELOPMENT

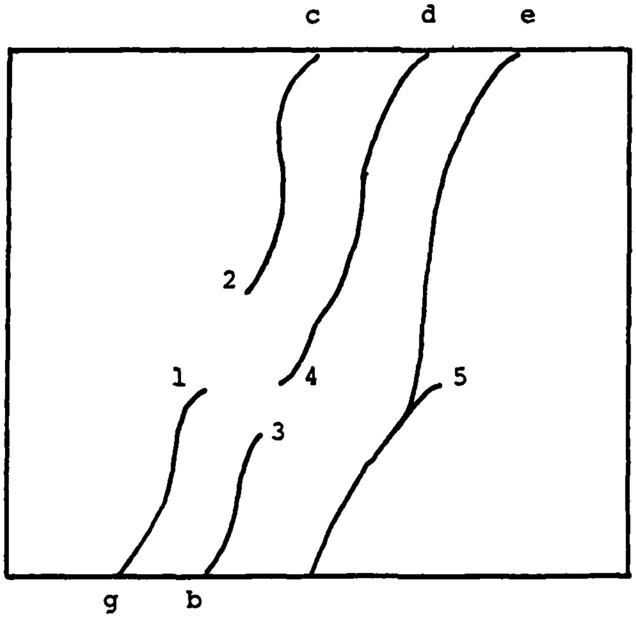


FIGURE 4.2: END POINT IDENTIFICATION

continued, contours d and h would next be connected. For the example used, the technique used for deleting the spur ending at point 5 could be quite simple because it is the only point remaining on the plot and the other end of the spur starts at a point with vectors in three different directions. However, the actual technique developed would have to be much more comprehensive. If, for example, the spur had occurred on the left side of the plot, it would have been the first point processed. Therefore, the technique must check each point as it is selected for processing to ensure that the line terminating at that point actually starts as a unique contour somewhere on the edge of the plot.

When automatically connecting a broken contour, the software should ensure that no intersections with another contour are created. A report would also be made listing all editing performed as well as any errors detected that cannot be resolved; to permit human review, resolution of any remaining ambiguities, and final acceptance.

After these techniques are developed, tested, and refined; more difficult types of contour problems can be addressed. An example that has been provided to ANDRULIS on an overlay is a contour plot that contains a ridge. Programatically, the ridge can be considered to be several contours that occupy the same space. Given this assumption, software can be developed to identify and tag ridges. ANDRULIS looks to the DMA for setting the priority on the types of contour editing of tagging problems that exist, and the need for automation in aiding the resolution of these problems.

4.2 AUTOMATED DLMS TAGGING

The experiments conducted in the area of tagging DLMS data utilized idealized data. The location of the centers of the label frames were very close to the features. The data for generating the label frames and the DLMS feature data were assumed to be present in two separate files. Neither one of these assumptions necessarily is representative of the currently used techniques. The DLMS overlay provided contained both the features and labels. Different colors were used on the overlay, so a color scanner could be used to generate the separate files. However, if the AGDS System is to be considered as the basic operating system for DLMS processing, data compatibility can be a problem. ANDRULIS suggests that consideration be given to the use of a separate overlay for labels. Use of an overlay for the feature number labels will eliminate the need for arrows; which are only useful for human readability and add to the problem of automated processing.

ANDRULIS recommends that the next step in development of an operational automatic feature identification and tagging technique should utilize real data, generated by the AGDS System as separate features and label files. This data should be supplied as both run length encoded and vectorized data for each set. The reason that the data should be processed in both raster and vectorized forms is so that the effectiveness and efficiency of processing can be compared between the two types of source data. It is possible that one form of the data would be most effective for locating labels and the other form would be more suitable for the actual association of labels and features.

The purpose of the experimental testing with AGDS data will be to determine the best techniques for setting distance tolerances for associations and resolving ambiguities. Since the header data contains information on feature type, computerized validation and ambiguity resolution can be tested using this information. The overall tagging capability should be designed to operate in a batch mode and create a temporary output file and report that can later be processed by an editor in an interactive environment. During the editing phase, the operator will resolve any remaining tagging errors and make final acceptance of the processing. The final data file will then be created.

SECTION 5
CONCEPTUAL SYSTEM DESIGN

5.1 GENERAL DESIGN CONCEPT

The general logical macro-flow for an integrated system to process cartographic material is presented in Figure 5.1. This representation does not consider the specific hardware/software interrelations or the form of intermediate or final data storage. Tape symbols are used to depict the intermediate files created that will be the output of one set of procedures and the input to another set of procedures. Some of the procedures depicted are being conducted on currently existing systems at the DMA facilities. Other procedures could possibly be developed by software implementation on current systems. Special purpose hardware should also be considered for implementation of new procedures. Development of the logical techniques should be completed before an implementation plan is started. The five sets of procedures are depicted as being independent, with only the transfer of data required between them. The first four are all batch processing and could operate as one job stream. The final procedure is an interactive one that does not appear suited as part of the job stream. It can be conceptually perceived that the batch procedures can be developed to the extent that the interactive procedure is not required. A more practical goal is to develop a system that identifies those data sets that need interactive review and automatically creates the completed files for all others. For purposes of system acceptance and quality control, human review should also be available for these "good files." Each of the five separate sets of procedures will be discussed in the following paragraphs.

5.2 LABEL DATA PREPARATION

This procedure set is labeled A in Figure 5.1. The purpose of the procedures is to create a temporary data set in machine-readable

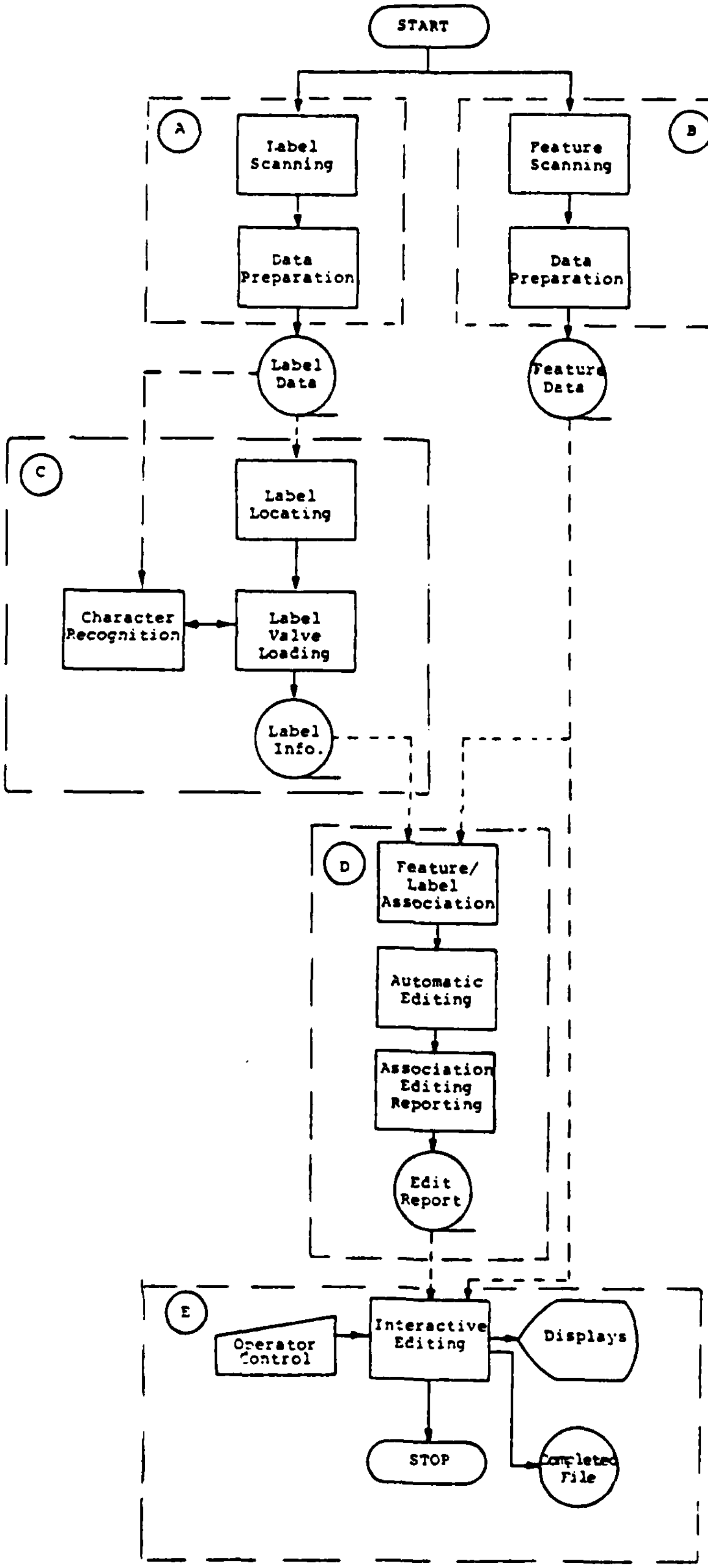


Figure 5.1: Conceptual System Design

format for locating the labels and obtaining the label values. The final format of the label data, vectorized or run-length encoded, has yet to be determined. The scanning of the label data manuscript could be done on either the AGDS raster scanner if the label data is on a separate overlay, or a color scanner if the labels are on the same manuscript as the features but in a different color. If vectorized data is required for either the character recognition capability or the label locating process, the AGDS vectorizing capability could be utilized. Therefore, procedure set A, although it is not now done, could basically be implemented on existing DMA systems. This could require the resolution of compatibility problems between the AGDS and color scanner.

5.3 FEATURE DATA PREPARATION

This set of procedures, labeled with a B in Figure 5.1, is concerned with the preparation of feature data for further processing. The term features is used here to encompass whatever type of cartographic data is being processed. All of the information presented in Section 5.2 would also apply to this set of procedures. This type of activity is currently being conducted by DMA, although some compatibility modifications might be required.

5.4 LABEL PROCESSING

The purpose of this section, depicted as C on the flow chart of Figure 5.1, is to obtain the location and value of the labels. Experiment two in Section 3 of this report addresses the overall logic of this set of procedures. For those types of cartographic data for which there is additional header information associated with each label, it is assumed that the header data will have been prepared in a separate machine-readable format. This information may be read at the label value loading step and incorporated into the label information file; or it may be introduced into the system at the beginning of the next set of procedures. To ANDRULIS' knowledge, none of the capabilities depicted in this step currently exist, although a character recognition capability is under development.

5.5 ASSOCIATION AND EDITING

Section D of Figure 5.1 depicts the key section of this conceptual system. The techniques for the identification of features and their association with labels were tested experimentally as explained in Section 3 of this report. The automated editing would be based on the techniques recommended for further development in Section 4 of this report. The output of this set of reports could be entirely a machine-readable file, or it could be a hardcopy report and a control file.

5.6 INTERACTIVE REVIEW AND EDITING

ANDRULIS envisions that the hardware/software of the AGDS editor's station could be utilized for this final set of procedures, *labeled with an E in Figure* 5.1. The feature data would be displayed as it currently is done. The edit report file would prompt the operator with any unresolved or inconsistant associations or data editing required. It would also have the capability to display under operator control, all associations sucessfully performed and all data corrections completed. The operator would have final control of accepting or modifying this information. When the operator is satisfied, the completed, digitized cartographic data set would be generated.

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.

DATE
ILMED
-8