

OPTICAL CHARACTER RECOGNITION
FOR
READING MACHINE APPLICATIONS

by

JON KAUFMANN CLEMENS

B. A., Goshen College
(1960)

B. S., Massachusetts Institute of Technology
(1963)

M. S., Massachusetts Institute of Technology
(1963)

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September, 1965

Signature of Author
Department of Electrical Engineering, August 23, 1965

Certified by _____ Thesis Supervisor

Accepted by _____
Chairman, Departmental Committee on Graduate Students

OPTICAL CHARACTER RECOGNITION
FOR
READING MACHINE APPLICATIONS

BY

JON KAUFMANN CLEMENS

Submitted to the Department of Electrical Engineering on August 23, 1965 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

Optical character recognition was investigated, from the viewpoint of specific use in a reading machine for the blind. It was found that existing optical character recognition techniques were unsuitable for this purpose, so that a different approach was required.

In the technique developed here, a stroke sequence for printed characters is generated by contour tracing the outside of the character by means of a digitally controlled flying spot scanner. In principle, this technique is based on the definition of lines which in turn define the character. The waveforms comprising the horizontal and vertical components of this trace are smoothed during the scan operation. The local maxima and minima of these smoothed waveforms are detected as well as the position at which they occur and are used to recognize the character. A non-linear smoothing technique is used which allows a large amount of smoothing while preserving the significant features of the character. The major noises of the printing process, that of style variations and ink run, are virtually eliminated by this smoothing technique.

Since the noise elimination is done in the scanning process, the recognition equipment can be very simple and therefore relatively inexpensive. It is estimated that the recognition equipment, including the scanner and associated equipment, can be built for a few thousand dollars.

The results of experimentation with this technique, including a page read from a popular magazine, are presented to demonstrate the validity of this technique for reading machine applications.

Thesis Supervisor: Samuel J. Mason
Title: Professor of Electrical Engineering

ACKNOWLEDGEMENTS

The author is sincerely thankful for the many ideas for investigation, the advice on methods of attack for the problems encountered, the inspiration, and the concern given to him by his advisor, Professor S. J. Mason. Working under Professor Mason has been both an enjoyable and valuable experience.

The author also wishes to thank his readers, Professor M. Eden and Professor H. Teager for their advice, encouragement and helpful guidance.

The author is also indebted to the members of the Cognitive Information Processing Group of the Research Laboratory of Electronics whose help, arguments, and encouragement were much appreciated.

The author also wishes to acknowledge the Research Laboratory of Electronics for providing the necessary facilities for research and the M.I.T. Computation Center for computer privileges.

TABLE OF CONTENTS

4.

ABSTRACT.....	2
ACKNOWLEDGEMENTS.....	3
LIST OF FIGURES.....	8
CHAPTER 1. Introduction.....	9
1.1 Motivation.....	9
1.2 Reading Machine Output Devices.....	9
1.3 Machines Which Do a Recoding of the Optical Signals.....	10
1.4 Machines Which Use a Sequence of Characters....	11
CHAPTER 2. A Personal Reading Machine.....	13
2.1 Introduction.....	13
2.2 A Reading Machine System.....	13
CHAPTER 3. Requirements on Optical Character Recognition (OCR) Equipment.....	17
3.1 Introduction.....	17
3.2 Speed Requirement.....	17
3.3 Error Rates.....	17
3.4 Numerics.....	19
3.5 Type Font Variability.....	19
3.6 Low Cost.....	20
3.7 OCR Requirements for Electronic Data Processing	21
3.8 Properties of the Printed Material to be Read..	22
CHAPTER 4. The OCR Equipment Problem.....	24
4.1 Present OCR Machines.....	24
4.2 Review of OCR Techniques.....	24

	5.	
4.3	Printing Noise.....	27
4.4	Evaluation of these Techniques for Reading Machine Applications.....	28
4.5	Basis of Technique Used.....	31
CHAPTER 5.	The Contour Tracing Technique.....	33
5.1	Captive Scan.....	33
5.2	The Scanner.....	33
5.3	The Digital Tracing Routine.....	34
5.4	Theorems and Properties Based on the Hexagon Tracing Routine.....	38
CHAPTER 6.	Smoothing the Trace.....	44
6.1	Introduction.....	44
6.2	Mouse-With-Tail Tracing.....	44
6.3	Radius of Curvature Momentum.....	46
6.4	Gear Backlash or Hysteresis Smoothing.....	47
CHAPTER 7.	Binary Description of Characters.....	52
7.1	Determination of the Smoothing Threshold.....	52
7.2	Generation of Binary Description of Characters.	53
7.3	Recognition of the Unknown Character.....	58
7.4	Significance of the Binary Description.....	60
CHAPTER 8.	Properties of this OCR Technique.....	64
8.1	Printing Noise Elimination.....	64
8.2	Character Size, Orientation, and Position.....	65
8.3	Low Bit Rate.....	67
8.4	Speed.....	67
8.5	No Feedback.....	68

8.6	Overlapping Characters Discriminated.....	68
8.7	Low Cost.....	69
8.8	Simple Tactile Display.....	70
CHAPTER 9. Experimental Data.....		71
9.1	Purpose of the Data.....	71
9.2	The Experimental Setup.....	71
9.3	The Learning Program.....	72
9.4	Multiplicity of Code Words.....	74
9.5	Confusion Among Letters.....	77
9.6	Tests on Standard Alphabets.....	79
9.7	Example of Machine Reading.....	84
9.8	Reading After Short Training.....	86
9.9	Validity of this Technique for Reading Machine Applications.....	92
CHAPTER 10. Areas for Further Research.....		95
10.1	Introduction.....	95
10.2	Test for Ascenders and Descenders.....	95
10.3	Eliminate Start Position.....	96
10.4	Test on Contour Size.....	97
10.5	Change of Areas for Coordinate Definition.....	97
10.6	Test for Leading Vertical.....	97
10.7	Test of Character Interior.....	98
10.8	Relative Positions of Neighboring Extrema.....	99
10.9	Tracing in Reverse Direction.....	99
10.10	Extrema Pairs Generation.....	101
10.11	Non-Orthogonal Directions.....	102

	7.
10.12 First Prototype.....	102
CHAPTER 11. Summary and Conclusions.....	104
BIBLIOGRAPHY.....	109
BIOGRAPHY.....	112

LIST OF FIGURES

8.

FIGURE	TITLE	
1.	A Personal Reading Machine System.....	15
2.	Hexagon Tracing.....	36
3.	Examples of Scanning Procedures.....	41
4.	Point Connectivity by Hexagon Tracing.....	43
5.	Hysteresis Smoothing.....	50
6.	Binary Description of Unknown Character.....	55
7.	Reconstruction of Unknown.....	62
8.	Significance of Coord Word.....	62
9.	Detected Extremum Position.....	62
10.	Letter "a" Found in the New York Times.....	76
11.	Confusion Among Letters.....	78
12.	Letters from 10 Different Alphabets.....	82
13.	Confusions from Time Magazine, June 11, 1965.	87
14.	Example of Reading Time Magazine.....	88
15.	Reading After Short Training.....	93
16.	Examples of Technique Variations.....	100

CHAPTER 1

Introduction

1.1 Motivation.

Optical character recognition (abbreviated OCR) is used throughout this paper to mean the process of classifying the optical signals obtained by scanning a page of printed characters into the same classes a human observer would class them. OCR machines have been used for some years in Electronic Data Processing. No personal reading machines for the blind, however, have been built which recognize the printed characters by machine. The reason for this is the cost of the equipment needed to do the recognition and the inability of present machines to recognize the many type styles which appear. The requirements on OCR equipment in Electronic Data Processing are quite different than those of a reading machine for the blind. It was felt that the special requirements for a sensory-aids reading machine system might lead to a different and more economical solution of the character recognition problem; one in which there is considerable interplay between the machine and the reader.

1.2 Reading Machine Output Devices.

In a reading machine for the blind the material must be presented to the reader in some manner other than optically. Several different types of output devices have

been studied experimentally. Since the optical character recognition device must supply one or more of these devices, a short summary of the various devices is in order. In theory, a reading machine for the blind can operate anywhere in the range between machines which do a one to one mapping or a recoding of the optical signals obtained from a printed page into other signals used for presentation and machines which classify these optical signals into character or word classes and present one member of this class. The existing machines, however, can be largely grouped into machines which do a recoding of the optical signals and machines which use a sequence of characters obtained by another source.

1.3 Machines Which Do a Recoding of the Optical Signals.

One class of reading machines merely transforms the optical signals into other signals for presentation.

In one of these machines a photocell array is placed over the material to be read. Each photocell controls one solenoid in a corresponding array and the solenoid is activated when the photocell detects black. The reader can now feel the letter shape and can recognize the printed letter. In a variation of this device some operation is performed on the output of the photocells before they are applied to the solenoid array. For example the point of

maximum curvature of line segments could be presented.

In another of the recoding machines an array of photocells activates an array of oscillators. This allows the operator to hear whether the photocells detect black or white. The Battelle Optophone⁽¹⁾ uses this technique. Reading rates higher than 30 words per minute are difficult to obtain with a machine of this type.

1.4 Machines Which Use a Sequence of Characters.

Another class of reading machines presents in some manner a sequence of letters supplied by some other device.

One possible output for a machine of this type is to code the letter into non-voiced tones, tactile signals, kinesthetic signals, or some combination of these. The operator then learns to associate the code with its corresponding letter. Reading rates of 60 words per minute or more have been obtained with machines of this type: see the Ph. D. theses of D. Troxel⁽²⁾ and R. W. Donaldson⁽³⁾.

Another output for a machine of this type is that of spoken letters. That is, the letter to be presented is spoken to the reader. If care is taken the time required to say a letter can be made relatively short. Metfessel⁽⁴⁾ has been able to obtain reading rates of over 100 words per minute with an output of this type.

The highest reading rates to date are obtained with

spoken words. The maximum is about 300 words per minute. Other techniques may supply higher reading rates in the future, but spoken word reading rates are by far the highest now. One possible way to obtain spoken words is to assemble the letters for one word and then look it up in a list of words which have been previously recorded. If the word is in the list it can be spoken, if not the letters of the word can be spoken. With a much more sophisticated machine it might be possible to generate spoken words from the sequence of letters with the use of a small word list and rules of the language. This problem is now under study in the Cognitive Information Processing Group of R.L.E.

CHAPTER 2

A Personal Reading Machine

2.1 Introduction.

For many kinds of reading it is desirable to have a personal reading machine. That is, a machine which is controlled by one operator so that he can go back over material, speed it up, slow it down, stop it, start it, etc. In other words the operator should be allowed to adjust all the controls. The machine of course does the recognition. By making use of the ability of the operator the requirements of the optical character recognition equipment can be reduced.

2.2 A Reading Machine System.

In Figure 1 is shown a possible reading machine system. The scanning equipment picks up the information from the printed page and converts it into electronic signals. The operator has control of the position of the scanner, its focus, its contrast, etc. The electronic signals from the scanner are now converted into signals which are appropriate to the recognition equipment. In general this amounts to a great deal of data reduction and is so labelled. There will also in general be adjustments for the quality of the printing and the style of type. Hopefully these will be just trimming adjustments

and not selection of the correct type font, since the blind reader does not, in general, know the type font.

In normal operation the output of the data reducing equipment will be connected by the switch to the recognition equipment. The recognition equipment either recognizes the letter or not depending on the distortion of the printed letter. If it is recognized it will be displayed. Assuming for example the display is spoken letters, the letter will be spoken to the operator. If it is not recognized an appropriate signal will be given to the operator. It is now up to the operator to decide whether he wants the distorted letter to be displayed by a recoding device such as a two dimensional pattern, or whether he wants the machine to go on.

As an example let us assume that the "e" in the word "the" is distorted beyond machine recognition. The machine could say t - h - beep. The operator would either understand the word or not. If he did not he would throw the switch and the pattern of the distorted "e" would be displayed as a tactile two-dimensional pattern, independent of how many correct letters had been read after this bad "e". The operator could now feel the letter and tell the machine to go on.

This is only one example of a possible machine system.

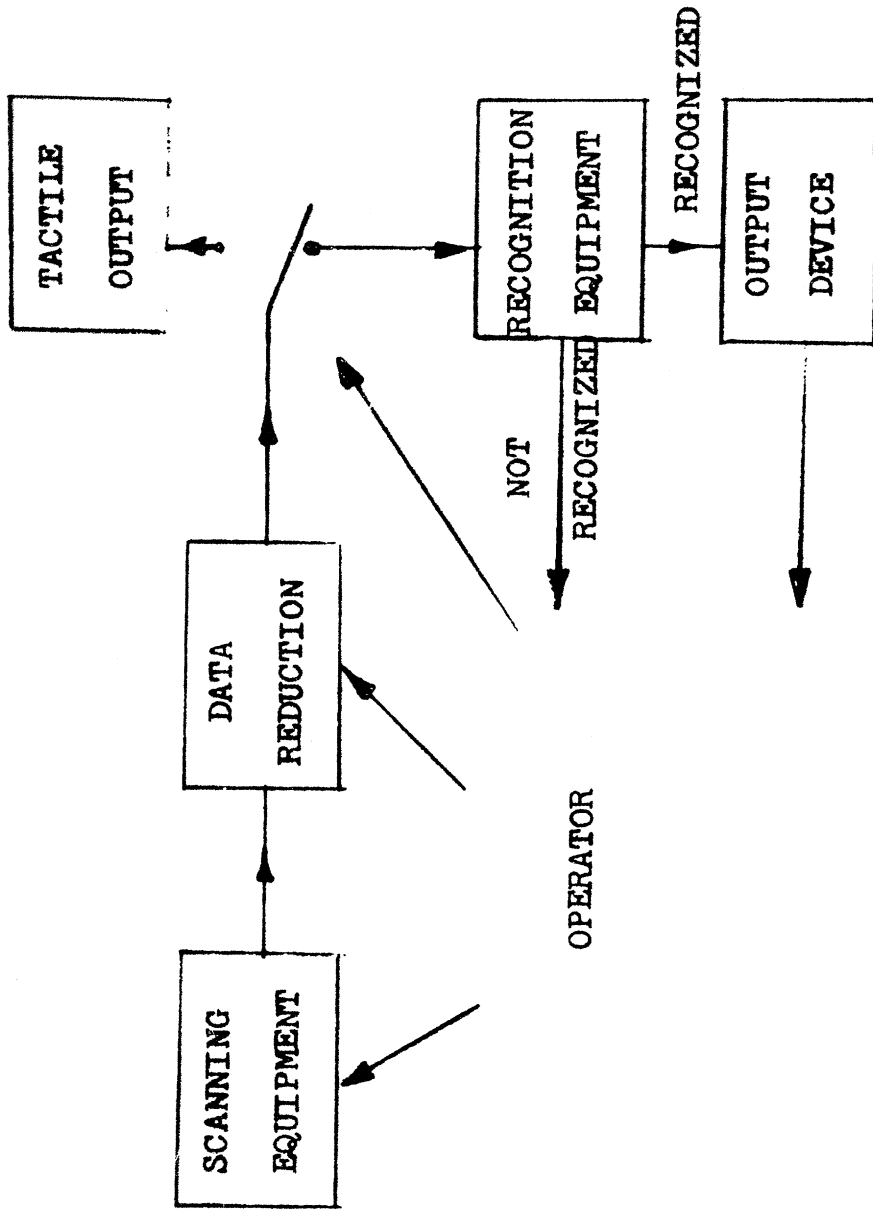


Figure 1. A Personal Reading Machine System.

Notice that there is no feedback from the recognition equipment to the scanner and data reduction equipment except through the operator. This would be a very valuable feature if a very complicated and expensive word speaking display were used, since it would allow the expensive machine to be time shared without losing the personal machine qualities.

This reading machine system is a combination of the recognizing and recoding machines mentioned in Chapter 1. Almost any combination of any of the types of these machines is a possibility. A good combination must be used for an effective inexpensive reading machine.

CHAPTER 3

Requirements on Optical Character Recognition (OCR) Equipment

3.1 Introduction.

When optical character recognition equipment is used in a reading machine system such as that described in Chapter 2, certain requirements are relaxed and some made more exacting in relation to conventional OCR machines. In this chapter the requirements of OCR in this reading machine system are given. They are then compared to the OCR requirements in Electronic Data Processing.

3.2 Speed Requirement.

As stated in Chapter 2, the highest possible reading rate today is about 300 words per minute. If an average of 5 letters per word is assumed this is equivalent to 25 letters per second. Of course if there are a lot of unrecognized letters due to distortion the reading rate will drop way down, but a reading machine should be able to operate at this rate given clean letters to recognize.

3.3 Error Rates.

The English language is very redundant. Shannon⁽⁵⁾ estimated that printed English text is approximately 75% redundant. That is to say if all sequences of letters were equiprobable the same information could be

transmitted with one-fourth as many letters. This implies that some errors can be tolerated without loss of information. Miller and Friedman⁽⁶⁾ have shown that the average person, given limited time to work, will not be able to correct passages perfectly if more than 10% of the characters are mutilated. Mutilation means, in this case, substitution of a character with another symbol. They also found that the job is most difficult if the mutilation consists of random substitutions with erroneous characters. The job was found to be easier if the position of substitution was marked and even easier if the letters were merely deleted and the position of deletion marked.

In a reading machine such as the one described in Chapter 2 with spoken letter output, the reading rate would be slowed down greatly if the reader had to resort often to the tactile two-dimensional display. It appears that this would not be necessary if the error rate of the OCR equipment was less than 10%. Of course the reading rate would increase as the error rate decreased.

The limit of the error rate is of course controlled by the error rate of the printed material itself. This is sometimes higher than 1%. The OCR equipment then should have an error rate lower than 10% and as close to 1% as possible. Possibly an error rate of 5% would

be sufficiently low.

3.4 Numerics.

There is no need to distinguish letters from numbers which are topologically similar. While reading prose the reader is able to sense when a character is a number and not a letter. This is portrayed by the fact that on many typewriters the small letter *l* and the number 1 are the same key. The numbers must be distinguished among themselves, of course, but the confusion between letters and numbers is not a serious problem.

3.5 Type Font Variability.

An important requirement of a personal reading machine is that it be able to read the various type styles which appear in magazines, newspapers, and books. This requires great versatility in the OCR equipment.

A machine that has to be changed greatly to match a given font is not a satisfactory machine. In general the blind reader does not know what the type style is and a complicated trial and error procedure to find the correct font would be undesirable. Requiring the operator to switch between Sans-Serif and Roman type styles would possibly be tolerable but even this could be confusing when titles are printed in one style and the text in the other.

A desirable reading machine would be one that has a reasonably high recognition rate independent of the type style and any adjustments that need to be made are just trimming adjustments. These adjustments would be similar to the adjustments of focus and contrast on a TV set. That is, it is not even necessary to know what the controls actually adjust in order to get good operation easily. The adjustments on a reading machine would be noise threshold adjustments to cope with printing quality and other adjustments of this nature.

3.6 Low Cost.

An important requirement in a personal reading machine is that of low cost. It is desirable to have the reading machine a personal one. This implies that the cost of the reading machine must be low enough so that the average blind person can afford it. A reading machine that costs about the price of a car is generally considered well priced.

The OCR equipment that exists today is much too expensive to qualify as a personal reading machine. Such machines might be applicable to library situations where books are converted to brail or to spoken words on tape. The cost of such a machine, however, makes its use for a personal machine impossible.

Future advancements in Electronic technology may very well make expensive techniques inexpensive. For this reason no technique should be eliminated from investigation on this point alone. When a technique for a machine to be used in the near future is searched for, however, the least expensive technique satisfying the requirements should be considered first.

3.7 OCR Requirements for Electronic Data Processing.

The requirements on the OCR equipment in reading machines for the blind differ widely from the requirements for the same equipment in Electronic Data Processing (EDP) applications. The requirements of a typical EDP application will be compared here to the requirements given above.

Speed requirements in EDP are often over 1000 characters per second. This is typically two orders of magnitude higher than that necessary in a reading machine for the blind.

Allowable error rates in EDP are very small. Unrecognizable reject rates, that is where a character is not recognized but is not confused with another letter, are on the order of one character per 2000 characters. Substitution error rates are on the order of one character per 100,000 characters. This compares to reading machine substitution error rates between one and ten characters per 100

characters.

In EDP there is a necessity to distinguish numbers from letters since data often is given with a mixture of the two.

In most EDP applications there is not a need for an ability to recognize an arbitrary typefont. Most of the documents are "turn around" documents. That is, they are typed, mailed out, and read on their return. In general the machines in EDP read only a few fonts and great cost is required to increase the number.

The requirement on cost in EDP is that the machine be able to pay for itself. This cost can be very high. A typical cost for a machine that reads several fonts would be between 100,000 dollars and 500,000 dollars. This is a factor of 100 times too large for use as a personal reading machine.

3.8 Properties of Printed Material to be Read.

The main use of a personal reading machine for the blind will be to read books, magazines, and newspapers. The OCR equipment can be designed, then, to the properties of this material. There are three basic properties.

One property is the fact that the letters were designed to be printed separately. This implies that in general the letters will be non-touching. This eliminates

certain printing used in advertising. A second property is that the letters are printed on a plain background. A third property is that the lines of the letter are single and continuous. Again advertising printing which used double lines or dashed lines is eliminated.

Although this simplifies the input to the reading machine, that amount of text which does not have these properties is minimal.

CHAPTER 4

The OCR Equipment Problem

4.1 Present OCR Machines.

The OCR machines which exist today were found unsuitable for reading machine applications. The reason for this stems from the far different requirements placed on machines used for Electronic Data Processing. There seem to be two main reasons for this unsuitability: Either the technique used to recognize the character was much too expensive or the machine was unable to read varied type styles. An attempt was made to make these machines suitable by reducing the requirements imposed by EDP. Again the attempts were unsuccessful because of two main reasons. In some cases even when the requirements on error rate and speed were reduced the technique used to recognize the characters remained basically unchanged and the cost remained high. In other cases the error rate deteriorated rapidly when differently designed fonts were introduced because the technique was designed for the peculiarities of one type font.

For these reasons it was necessary to design a technique suitable for a reading machine.

4.2 Review of OCR Techniques.

In order to discuss the various types of OCR techniques,

the techniques have been separated into four classes. Admittedly these classes are somewhat arbitrary, therefore some schemes may cover a combination of classes and others may fall in between classes.

The first class is that of template or mask matching. Although there are many ways of performing this operation, they are equivalent to making a transparency of a character shape and comparing it optically to the unknown. This transparency can be made to take into account the statistical variations of the character. The unknown character shape is generally picked up optically but can be compared either optically as in one of the RCA machines⁽⁷⁾ or electronically as in the Philco machine.⁽⁸⁾

The second class of techniques are those that search for local features all over the area occupied by the unknown character. In this class, a feature to be measured is selected by the operator. Then in some manner a small sub-template is made of this feature and it is compared all over the character area in some orderly fashion. Many different features can be chosen and the results of all these tests are combined and used to decide the identity of the unknown. It is important to note that a person chooses the features and that the search for position of the features is not controlled by the unknown character at all. Two examples of this class are the search for

vertical and horizontal lines, Farrington⁽⁹⁾; and looking for line segments and their intersections, Bomba⁽¹⁰⁾.

The third class of techniques is that of captive scan or feature searches which are controlled by the unknown letter itself. In this way the analysis of the character follows the lines of the character. Recognition is based on the length and curvature of lines, the intersections of lines, or the stroke like sequence of the trace itself. Examples of this technique are those of the watch-bird scheme used by Rabinow⁽¹¹⁾ and the contour tracing used by Greanias⁽¹²⁾ and by Bradshaw⁽¹³⁾.

The fourth class of techniques are those which make random measurements on the unknown character and compare them to a learned set. The word random is intended to mean that the choice of what features to look for is left to chance. The machine itself decides the best decision procedure under the restrictions of the machine wiring. An example of this general technique is that of Bledsoe and Browning⁽¹⁴⁾ where 75 point pairs are picked at random and are compared to the same measurements on a known list of characters. The perceptron⁽¹⁵⁾ is the limit of this group. In this machine there is as little human intervention as possible.

There are many equivalences between classes. For

example all four classes can be considered template matching if complicated templates can be used. In fact, all the character recognition schemes do the same job, when given an array of black and white spots (quantized output of a scanner) they give a yes or no answer stating whether the input was a certain character or not. In other words, they are all a Boolean operation on the input space. The difference between schemes is the way this Boolean operation is factored. This can be done in one step or many steps. A good scheme must eliminate the printing noise while maintaining character identity and do it efficiently.

4.3 Printing Noise.

In the character recognition process any phenomenon which adds variability to the unknown to be recognized is called printing noise. The major printing noises are: (1) Type style variations; (2) Black areas where white should be and white areas where black should be or "salt and pepper noise"; (3) Touching letters; (4) Broken character lines; (5) Rotated characters; (6) Displaced characters; (7) Line thickness variations.

By far the most common sources of noise to be encountered by a machine reading books, magazines, and newspapers are type style variations and line thickness

variations. All other noises occur very rarely. Possibly the next most common noise is that of displaced characters especially in newspaper text. It was surprising to find how rarely "salt and pepper" noise is found. A lot of machines are designed to eliminate "salt and pepper" noise simply because it is the easiest to model. Most of the other noises are dependent mainly on the printed character itself and are thereby almost impossible to model.

A reading machine of the type under discussion should be designed against the most common noises. The error rates caused by the other noises will be much below that required.

4.4 Evaluation of these Techniques for Reading Machine Applications.

The most successful OCR machines for reading varied fonts today are those of class one. The concept of the template, however, is basically one of a single font. For this reason many templates of each letter have to be stored or a sophisticated statistical test must be made. Each unknown must be compared to all the stored unknowns. For these reasons a varied font machine of this kind is very expensive. The existing machines cost a factor of 100 too much for use in a personal reading machine.

The techniques of class two are used mainly for specially designed fonts. In general this class of techniques has possibilities of being very versatile with respect to type style. Unfortunately they also require a great deal of logic and therefore are very expensive. It is possible to use these techniques simply and inexpensively with specially designed fonts because the exact position of the property searched for is known in advance. The cost of these techniques for varied font inputs is prohibitive.

The techniques of class four have produced relatively few working machines. The initial expense of a machine based on this class is very high because of all the machine learning required. Final machines of this class would not be build with the learning capabilities, but would be wired permanently to operate like the learned machine.

The techniques of this class neglect the fact that the characters were designed by humans to be recognized by humans. They do not use the knowledge we have about how humans recognize the characters. For this reason these machines are sensitive to noises humans reject well: those of varied type styles, varying line widths, and "salt and pepper noise".

The techniques of the third class show the most promise in terms of cost. This is true basically because the process of tracing the unknown character simplifies the scanning procedure, since there is a lot of area which is never scanned. In addition, it eliminates the printing noise of line thickness variations, since continuity, not thickness, of the line is important. The printing noises of broken lines and touching letters would be serious to this class. They are, however, minor noises and there are techniques to reduce these noises even further.

M. Eden⁽¹⁶⁾ and P. Mermelstein⁽¹⁷⁾ based the recognition of handwritten characters on the character strokes and the information about how the strokes were generated. Although they were not completely successful in the sense of complete recognition, their experiments did show that the line strokes of the characters carried important information needed to recognize characters.

Recently H. Teager delivered a paper⁽¹⁸⁾ in which a basis for measurements on patterns was presented. In his paper examples are given to support his views and a technique for the classification of handwritten characters with temporal information is given.

Very briefly, Teager shows that a pattern is composed

of a hierarchy of lines which are defined on the basis of the visual signal. This hierarchy of lines is defined by successive viewings of the pattern from low resolution to high resolution. The lines formed from the low resolution pattern are called figural lines. He then shows that the eye looks for points of interest along the lines of high resolution relative to the figural lines. Among the points of interest along the line, also called extrema, are its end points and its maxima and minima relative to the figural line. These extrema and their relative positions are invariants of the pattern.

According to Teager the figural lines of the characters of the Roman alphabet generally coincide with the absolute orthogonal directions. Teager made use of this fact in the design of the machine for the classification of handwritten characters. The basis for the classification of these characters is made on the points of interest along the lines forming the character. The major points of interest were those of line terminations and the maxima and minima of the line relative to absolute orthogonal directions. This method of classification was very successful, showing virtually no misclassifications.

4.5 Basis of Technique Used.

Machine printed characters, although generally more

regular, differ greatly from handwritten characters in three ways. First, there is no temporal information associated with the strokes, the character comes all at once. Secondly, the line width relative to the character size is generally much larger than that of handwritten characters. Thirdly, the amount of printing noise changes the line positions a significant distance relative to the character size.

Nevertheless, it was felt that the approach used in the techniques of class three could lead to a simple description of the character if line could be defined. In the Roman alphabet the outside contour of the character describes almost all the characters. For this reason then, line was defined by tracing the outside contour of the character to be classified. This trace was then smoothed and the local maxima and minima of this trace were used to classify the character. It will be shown that this procedure leads to a good engineering approximation of the desired information.

CHAPTER 5

The Contour Tracing Technique

5.1 Captive Scan.

In the technique used for character recognition in this thesis the strokes or the lines which describe the letter are traced using a captive scan. That is, the unknown letter itself guides the trace examining the letter. Assume for example that a single line of characters appears in the area to be investigated. A column scan is generated which scans from bottom to top and left to right until the letter is contacted. The scan then changes its mode and traces around the outside contour of the letter. The maxima and minima of the horizontal and vertical deflection voltages needed to guide the trace around the letter contour are recorded and coded in such a way that they can be used to recognize the unknown character being traced.

5.2 The Scanner.

The optical system used to pick up the information on the printed page is based on a flying spot scanner. The spot generated on the screen of a cathode-ray tube is focused onto the printed page. The light reflected from the printed page is measured with a photo-multiplier tube. If the reflected light exceeds a certain threshold it is

called white, if not it is called black. The position of the spot on the cathode-ray tube, and therefore on the page, is controlled by the horizontal and vertical deflection voltages. These voltages are digital voltages in which the smallest step change corresponds to a position step change very small compared to the character size.

In the simulation of the optical system a transparency of the page of printed material is scanned by a scanner and stored in array form in the memory of the TX-0 computer. The tracing routine was then performed in the stored array. Since the deflection voltages of the scanner are assumed to be digital, the storing of the array is a minor step.

5.3 The Digital Tracing Routine.

The object of this digital tracing routine is to trace around the outside contour of a character which is described by an array of regularly spaced black and white points. The description of the tracing routine will apply to any array consisting of at least two connected points (connectivity will be defined later) surrounded by a border of white points.

Many different routines could be used to perform the desired operation. The one described here, however, has some worthwhile properties and will thereby be described

in some detail.

The trace is generated using local operations. This implies that a long trace is no more complicated than a short trace and no additional memory space or extra logic is needed. In addition, only the coordinates of the starting point need be remembered to determine when the trace is complete.

The local operation which generates the trace is based on the hexagon shown in Figure 2a.

The hexagon consists of a center and six nearest neighbors (NN). The tracing procedure is done in a "leap-frog" fashion. Figure 2b shows a typical starting position, the lowest black spot in a non-white column. A search is started arbitrarily at NN 0 for the first point only and each nearest neighbor is tested in increasing numerical order modulo 6 until the first black nearest neighbor is found. The hexagon is then displaced so that its center is placed at the position of this just found black spot, i.e. the spot labeled 1 in Figure 2b. The new position is shown in Figure 2c.

The search for a black nearest neighbor starts at the nearest neighbor one higher in numerical order modulo 6 than the nearest neighbor which was the last hexagon center, i.e. the search will now begin at NN 5 in Figure 2c.

NN 2	NN 3	
NN 1	C	NN 4
	NN 0	NN 5

Figure 2a.

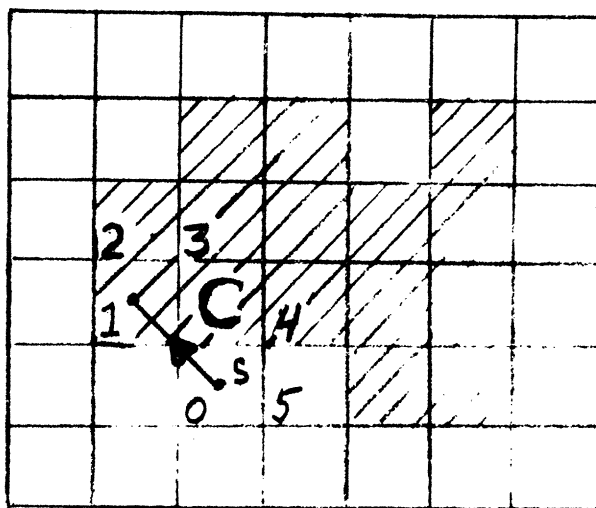


Figure 2b.

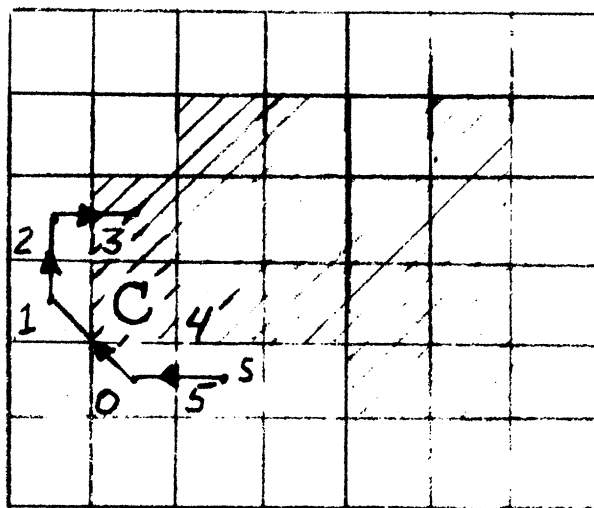


Figure 2c.

Figure 2. Hexagon Tracing.

This "leap-frog" procedure is continued until the starting point is again reached.

On the basis of this tracing routine connectivity can be defined.

Definition.

Two points A and B are connected if a hexagon centered at either point has the other point as a nearest neighbor.

Due to the properties of the hexagon tracing routine a series of single points lying on a 45 degree line are not connected whereas points lying on a 135 degree line are connected. For this reason it is necessary to insure that the thinnest line on any character be at least two points across. This will insure no breaks in the letter strokes.

The basic virtue of this tracing routine is that the logic needed to generate it is very simple. Suppose that the coordinates of the center of the hexagon are px, py . It is now desired to find the coordinates of $NN\ k+1$ when the coordinates of $NN\ k$ are known, k is an integer modulo 6.

Let the coordinates of $NN\ k$ be hx_k, hy_k , and the coordinates of $NN\ k+1$ be hx_{k+1}, hy_{k+1} , then:

$$hx_{k+1} = hx_k + hy_k - py$$

$$hy_{k+1} = px + py - hx_k$$

This is a very simple operation and needs very little logic. Many other tracing techniques require a stored list of instructions to designate the next move from a given initial arrangement.

This tracing routine, which will be called hexagon tracing, has some helpful properties which will be pointed out in conjunction with some theorems based on this routine.

5.4 Theorems and Properties Based on the Hexagon Tracing Routine.

Theorem 1.

Given two points in an array A and B such that A is the center of a hexagon and B is one of its nearest neighbors, then there exists a hexagon for which B is the center and A is one of its nearest neighbors.

This theorem is easily proved by enumeration. It is needed to assure that the "leap-frog" nature of the hexagon tracing is valid. It was already alluded to in the description of the tracing routine.

Theorem 2.

Given the hexagon tracing as previously described, the first point tested after a change of the center of

the hexagon will be a white point.

This theorem will be proved by contradiction. Let us consider three points A, B, and C, such that B is the present center of the hexagon, A is the previous hexagon center and C is the first nearest neighbor to be tested. Then the three points must be in one of the following 6 arrangements.

C	A	C	A	B	A	B	C	B
A	B	B	B	C	C	C	A	A

If C is black there is a contradiction. Since C is an earlier nearest neighbor than B to a hexagon centered at A, B could not be the hexagon center following a hexagon center at A. This completes the proof.

Theorem 2 enables the tracing to be faster since the first nearest neighbor need not be tested. It has greater value, however, since it allows the following theorem to be proved.

Theorem 3.

Consider an array of black and white points and a trace obtained from this array in the manner described previously. If the matrix is modified by changing to white all black points which have 6 white or 6 black nearest neighbors, a trace generated in the same manner on this modified array will be identical to the original trace.

If a point has 6 white nearest neighbors it cannot

be on a traced path, since the trace progresses from nearest neighbor to nearest neighbor. If a point has 6 black nearest neighbors it cannot be on a traced path by Theorem 2.

The operation described in Theorem 3 is that of edge extraction. Figure 3a shows a sequence of letters and the same letters after edge extraction. Note that black and white have been inverted in the picture. This edge extraction allows investigation of open sections in the interior of the letter. If the letter is contour traced and the contour is erased then the interior can be examined. This is shown in Figure 3b.

In a tracing routine of this nature it is an important feature to be able to detect the completion of the trace by testing each point's coordinates as they are found with those of the starting point. The following theorem guarantees that the hexagon tracing routine will always return to the starting point after tracing the unknown character exactly once.

Theorem 4.

If the hexagon tracing routine is started at the lowest black point of a non-white column, then the trace will always return to this point after tracing the outside contour of the connected black points of which this point

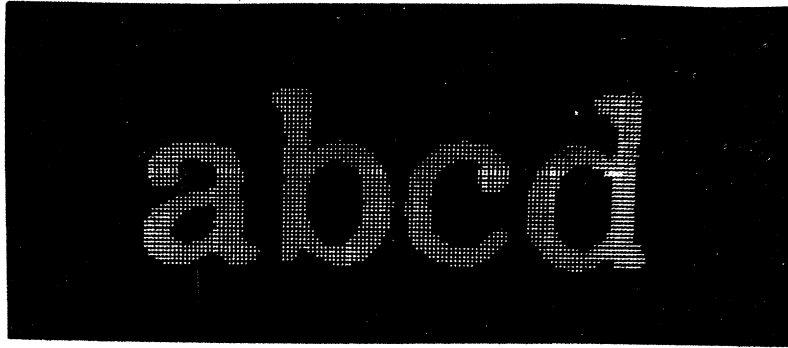


Figure 3a. Original Characters.

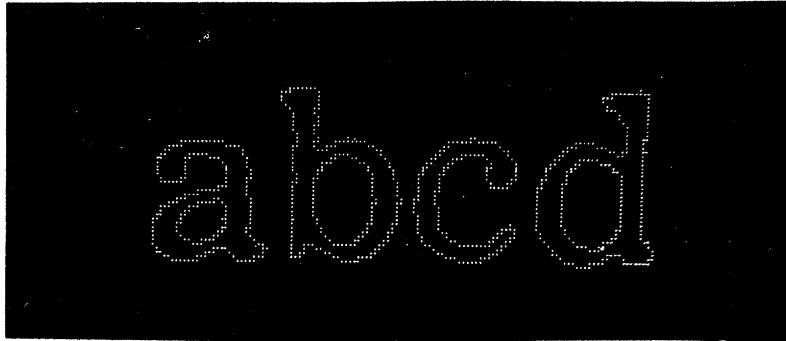


Figure 3b. Characters After Hexagonal Edge Extraction.

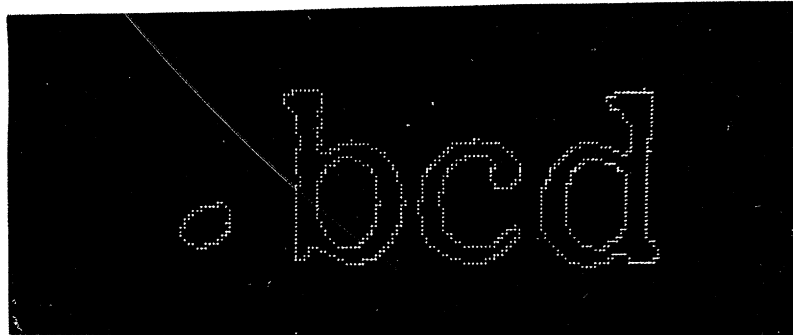


Figure 3c. Edge Extracted Characters After One Tracing With Erasure.

Figure 3. Examples of Scanning Procedures.

is a member exactly once.

In Figure 4a is shown a black point which is assumed to be on a trace. From this point arrows are drawn to every point to which the trace could possibly move in the next step. The arrows are drawn curved clockwise to indicate the clockwise "leap-frog" nature of the hexagon tracing.

Figure 4b shows an array of black and white points to be traced. If every black point is assumed to be the center of a hexagon and an arrow is drawn to each of its black nearest neighbors in the manner of Figure 4a, the result is Figure 4c. This operation breaks the points up into distinct figures. Each figure has a distinct outside. It can be seen from this construction and the fact that in the hexagon tracing routine the first possible path in the clockwise sense is taken, that the starting point will be reached after tracing the contour exactly once.

This theorem can be formally proved by using the fact that each traced point must have a white nearest neighbor which lies outside the traced contour. A formal proof will not be given here, however.

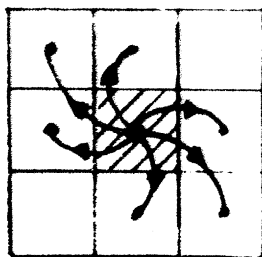


Figure 4a.

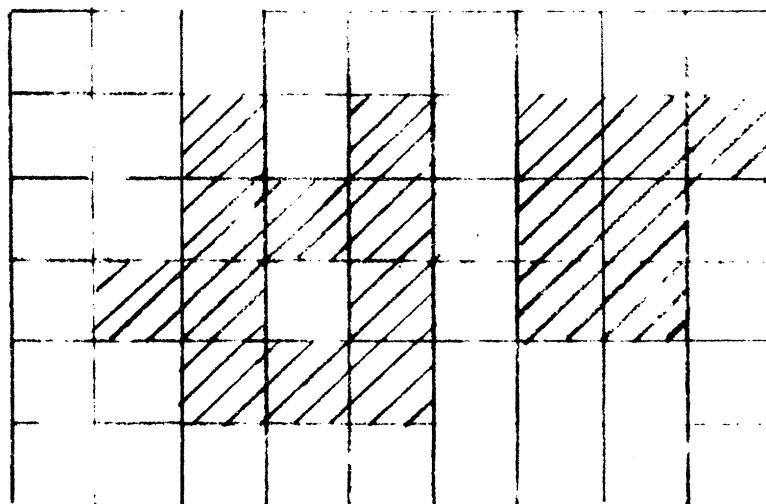


Figure 4b.

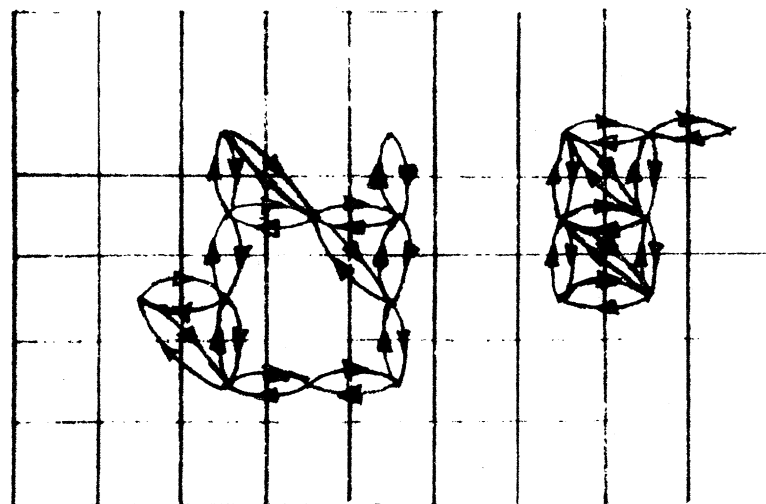


Figure 4c.

Figure 4. Point Connectivity by Hexagon Tracing.

CHAPTER 6

Smoothing the Trace

6.1 Introduction.

It is obvious that some type of smoothing is necessary while tracing the letter contour. The edge of a letter is generally quite rough due to ink run in the printing process. If no smoothing is done, there will be spurious maxima and minima on the curves describing the deflection voltages necessary for tracing the letter. Many different types of smoothing were considered. Two of these smoothing techniques, which looked promising at first but were finally unsatisfactory, were investigated in great detail. They will be described briefly here. Another type of smoothing which was found very successful will be described in detail.

6.2 Mouse-With-Tail Tracing.

One smoothing routine which was tried is analogous to a mouse with its tail. A square array of points 3×3 or 4×4 was defined which is analogous to the mouse. Another single point was defined a fixed distance from the mouse which is analogous to the point of the mouse's tail. Assume for example that the point of the mouse's tail is on the contour of the letter. The mouse is then adjusted until it is half black and half white or as

close to that as possible always remaining a fixed distance from the tail point. A direction is now defined from the tail point to the center of the mouse. A step is then taken in this direction moving both the mouse and its tail. Notice that this may force the tail point into the black area or into the white area. This is the smoothing property of the Mouse-With-Tail tracing. The amount of smoothing depends on the length of the tail.

This smoothing routine was found to be unsatisfactory because of the inconsistency of requirements in the length of the tail. If the tail was too short the amount of smoothing was not enough and the spurious maxima and minima persisted. If the tail was made long enough to get the correct smoothing the mouse would jump the gap between letters and sometimes get trapped in the interior of a letter and never get out. Another undesirable feature is caused by the integrating nature of the tracing routine. That is, the amount of deviation of the smoothed curve due to a feature such as the cross bar on a small t depended on the thickness of the feature itself. That is, a shorter but thicker black area, like printing noise, causes as much deviation as the longer but thinner feature if the areas are approximately equivalent.

6.3 Radius of Curvature Momentum.

Another tracing and smoothing technique which was experimented with can be called radius of curvature momentum. Very briefly, this method again used a 3 x 3 or 4 x 4 array of points to sense the edge of the character. The value of the sensing array did not change the trace directly, however. The trace proceeded in small increments around the character. The last three traced points were used to determine a center of curvature. With the new trace point situated on the letter contour a new center of curvature could be found using the last two previous points. The actual center of curvature was only allowed to move a fraction of the distance to this new center, however, with the new point placed this final circle. The label radius of curvature momentum is now obvious.

The net result was that the trace tended to stay in the curve determined by the last few values of the trace and smoothing occurred. Unfortunately there were many problems associated with this technique. Two of the major problems were that if enough momentum was imposed to perform the smoothing function adequately the trace would crash through parts of the letter where there was a thin line at a sharp angle. In addition the trace seemed to get trapped easily. Once the trace got into the center

of the capital "G" it would stay there indefinitely. The reason for this is obvious. These reasons made this technique unsatisfactory.

6.4 Gear Backlash or Hysteresis Smoothing.

The tracing and smoothing technique which showed success is analogous to mouse-with-tail tracing in that it has one point that traces the curve and the smoothing is done afterward. The smoothing is done with a non-linear operation and has distinct advantages over the integration smoothing of the mouse-with-tail technique and other linear operations such as two-dimensional exponential smoothing. These advantages will become clear in the description of the smoothing operation.

The smoothing procedure is similar to two gears which are meshed, but have some space between their teeth. This space permits the drive gear to turn back and forth without moving the load gear if the amplitude of the motion is smaller than the space between the teeth of the meshed gears. It can be said that the load gear has a motion which is a smoothed version of the drive gear motion.

It is important to note here that the smoothing operation is intended to eliminate the spurious maxima and minima of the unsmoothed curve. For this reason it

is unimportant whether the smoothed curve has sharp curves in it or not as long as the major maxima and minima are detected and well defined and the minor maxima and minima are eliminated.

The non-linear operation of this smoothing technique can be described with the electrical analogue of the gear backlash phenomenon. This electrical circuit is shown in Figure 5a. Ideal elements are assumed. If it is assumed that $F(x)$ has just increased a value larger than $2E$ it is clear that diode D_2 must be closed and diode D_1 open, therefore $G(x) = F(x) - E$. If $F(x)$ reaches some value M_a and starts to decrease, D_2 will open and $G(x)$ will remain unchanged until $F(x)$ reaches $M_a - 2E$. At this point D_1 will close and as $F(x)$ decreases $G(x) = F(x) + E$. Just the reverse happens at a minimum. If $F(x)$ reaches a low value M_n and starts to increase, D_1 will open and $G(x)$ will remain unchanged until $F(x)$ reaches $M_n + 2E$. At this point D_2 will close and $G(x) = F(x) - E$.

Note that when $F(x)$ changes so that both diodes are opened no changes in $F(x)$ will change $G(x)$ unless $F(x)$ moves so as to close one of the diodes.

It is very easy to detect a maximum in this way. If the slope of $G(x)$ changes sign a maxima or minima

is recorded. If the slope of $G(x)$ is positive and then becomes zero as $F(x)$ decreases but then changes back again to a positive value as $F(x)$ increases again no maximum is recorded. This eliminates the minor maxima and minima.

Figure 5b shows a curve $F(x)$ and its smoothed version level $G(x)$. The dotted curves are fictitious, showing E units above and below $F(x)$. Initially $F(x)$ is increasing and $G(x) = F(x) - E$. When $F(x)$ reaches a maximum and decreases $G(x)$ remains unchanged until $F(x)$ falls $2E$ below the maximum, then $G(x) = F(x) + E$. When $F(x)$ reaches a minimum and starts increasing $G(x)$ remains unchanged until $F(x)$ increases $2E$ above the minimum, then $G(x) = F(x) - E$. When $F(x)$ reaches the second maximum and begins to decrease $G(x)$ remains unchanged. In this case, however, $F(x)$ does not decrease $2E$ below the maximum. As $F(x)$ increases past the maximum then $G(x) = F(x) - E$ again. The maxima and minima are detected only when the slope of $G(x)$ changes sign. The major maximum and minimum are detected and the minor ones are eliminated.

The electronic circuit used here to describe the smoothing technique is only one way that this type of smoothing can be obtained. Many other arrangements can

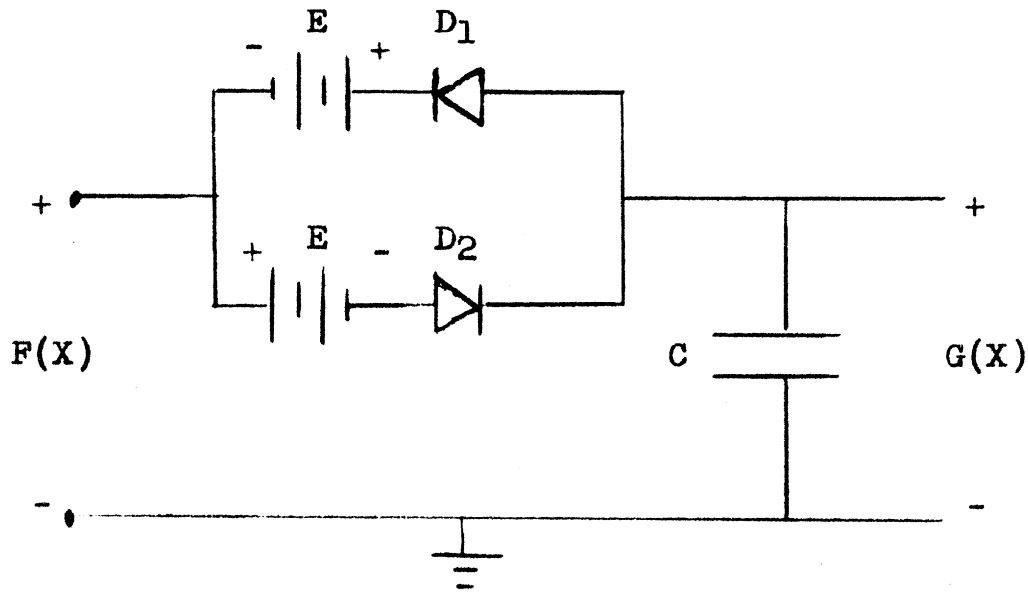


Figure 5a.

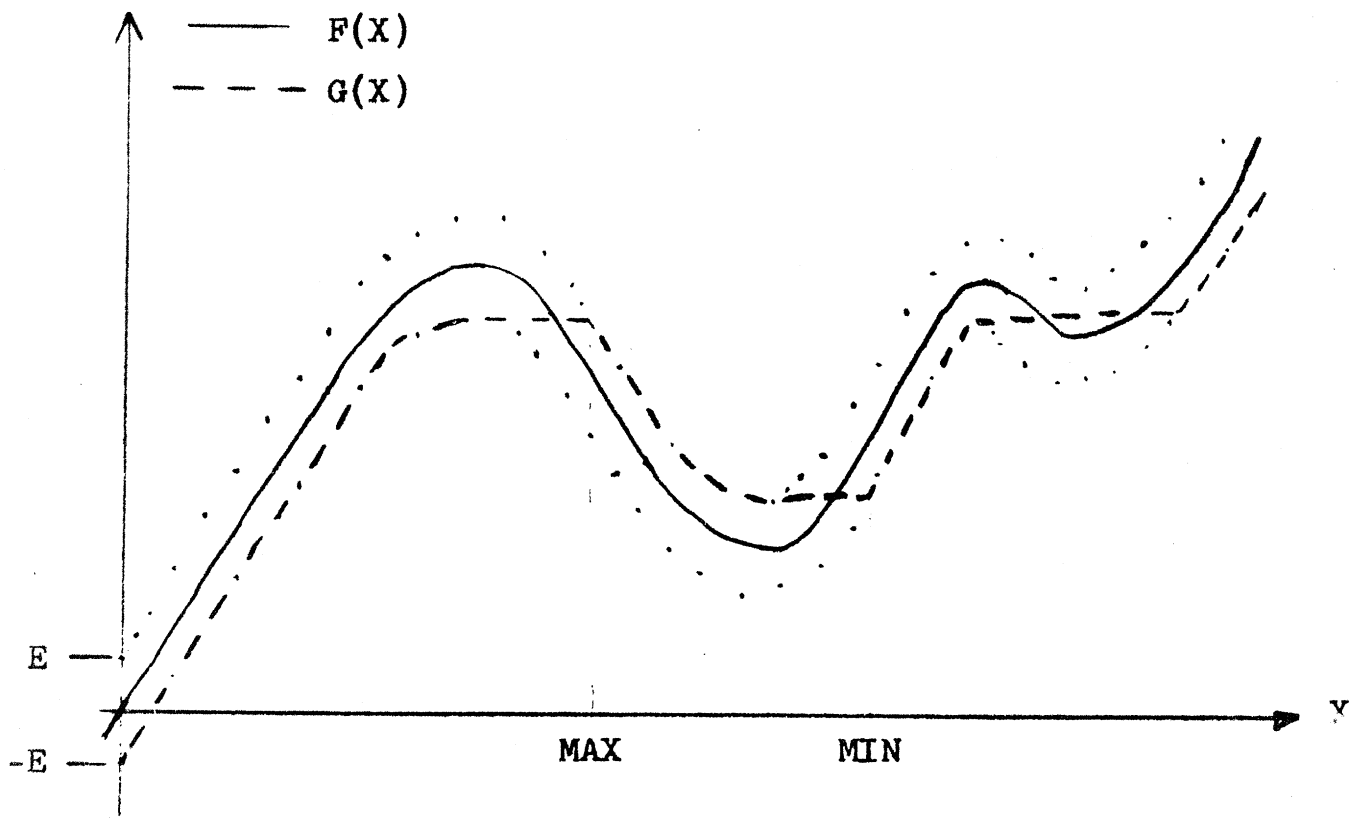


Figure 5b.

Figure 5. Hysteresis Smoothing.

supply the necessary hysteresis to do this type of smoothing. In the simulation the hysteresis is obtained by storing the coordinates of the highest point reached and comparing the coordinates of each new point with this value minus the threshold.

This smoothing technique has several advantages over the linear smoothing techniques. Unlike the integrating nature of the linear technique this smoothing depends on amplitude and not on area. This implies that narrow but significant features will not be confused with broad low amplitude noises such as ink run. In addition all motions no matter how frequent are completely eliminated if their amplitude is less than the threshold. That is, the jagged edge due to ink run is not smoothed somewhat but eliminated. Finally, this smoothing technique can be easily implemented digitally or with an analogue circuit.

CHAPTER 7

Binary Description of Characters

7.1 Determination of the Smoothing Threshold.

The recognition of an unknown character is based on the description of the horizontal and vertical waveforms generated while tracing the outside contour of the character. There two waveforms are described in terms of their local maxima and minima. The threshold used in the smoothing operation, that is the value $2E$ used in Chapter 6, should be large enough to eliminate spurious maxima and minima due to ink run in the printing process and other types of printing noise. In fact the threshold should be as large as it possibly can be and still retain the significant maxima and minima that distinguish the letter. A large threshold will eliminate the variants introduced by the type of serifs used and the style of the letter but it must not affect the significant features distinguishing the character.

It was a significant result of this thesis to discover experimentally that the smoothing threshold could be one-fourth the letter extent and larger. This was very surprising at first, but seems very reasonable on second thought. This large threshold implies that significant features of the letters are at least one-

fourth the letter extent in either the horizontal or vertical direction. For example, the feature that distinguishes the lower case "h" from the lower case "n" is at least one fourth the height of the "h". This large threshold eliminates virtually all of the noise introduced by ink run in the printing process and in fact almost all the variations introduced by the style of the letters as well as the style and placement of serifs.

7.2 Generation of Binary Description of Characters.

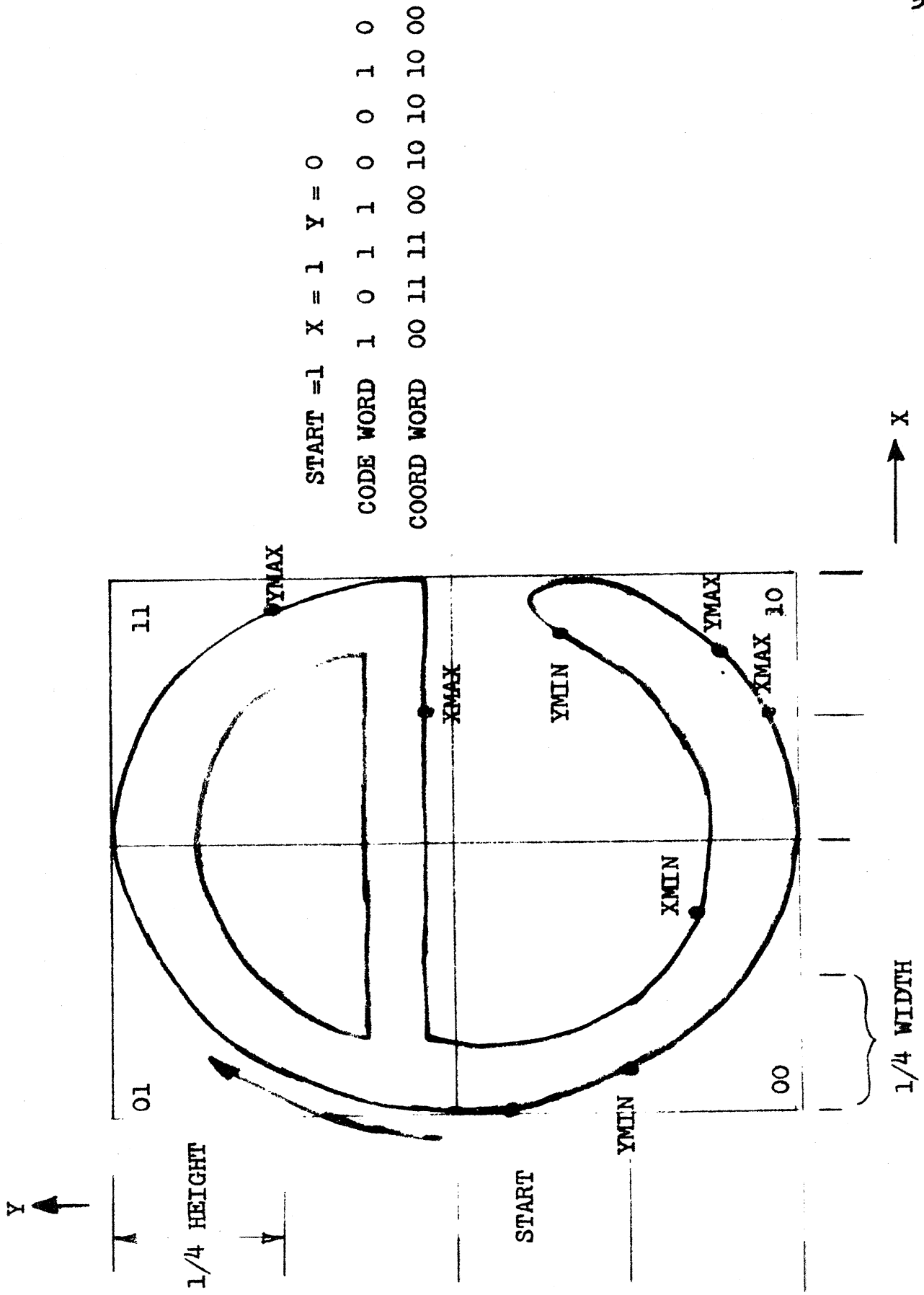
There are three separate binary words used in the recognition of the letters, the height-to-width ratio of the letter, the sequence of the maxima and minima, and the coordinates of the maxima and minima.

The generation of these binary words will be explained with the use of an example. Figure 6 shows a letter to be scanned and recognized. The search is started by a column scan, bottom to top, which moves from left to right until the letter is contacted. This will occur at the point labelled start in Figure 6. The captive scan described in Chapter 5 is then initiated and the outside contour of the letter is traced. During this trace only two pieces of information are recorded, the letter extent in the vertical or y direction and the letter extent in

the horizontal or x direction. From this information the height-to-width ratio is calculated. This information is also used to set the actual values of the smoothing thresholds. The waveforms occurring in the horizontal and vertical directions during the contour trace are considered independently. In this example the threshold for the vertical or y waveform is set to one-fourth the vertical extent and the threshold for the horizontal or x direction is set to one-fourth the horizontal extent. These thresholds are in general different as can be seen in Figure 6.

The trace routine is then initiated for the second time. During this trace the maxima and minima in the x and y directions are detected. Notice that it is not necessary to draw the smoothed curve as described in Chapter 6 to determine the position at which an extremum is detected. The extremum will always be detected when the trace moves one threshold unit from the extremum in the coordinate direction in which the extremum occurred. The next paragraph describes the detection of the extrema for the letter in Figure 6.

When the second trace is initiated both x and y increase until the top of the letter in Figure 6 is reached. At this point y starts to decrease and x continues



START = 1 X = 1 Y = 0
 CODE WORD 1 0 1 1 0 0 1 0
 COORD WORD 00 11 11 00 10 10 10 00

Figure 6. Binary Description of Unknown Character.

to increase. When y decreases the amount of the threshold, a y maximum is detected and is so labelled on the letter. As the trace continues y continues to decrease and x reaches a maximum and then decreases. When x decreases one-fourth the letter width a maximum is detected and is so labelled. As the trace continues again x reaches a minimum and increases. Notice that this minimum is detected at one-fourth the distance from the minimum and not one-fourth the distance from the edge of the letter. The trace continues detecting a y minimum, a y maximum, an x maximum and a y minimum before completing the trace.

The sequence of maxima and minima can be coded into a binary word, which will be called the code word. Each extremum of the trace can be determined non-ambiguously with one bit of information if it is assumed that the starting position is at a y minimum as well as an x minimum. The first extremum encountered on the trace must then be either a y maximum or an x maximum. This is a binary choice and can be designated with one bit of information. Each extremum can be designated by one bit since it is always known what the past extremum was in each coordinate direction. That is, although it may appear that there is a four way choice of the next extremum along

a trace there is in fact only two choices because of the history of extrema along the trace.

The code word is generated by recording a one for the start position, a one for an x extremum and a zero for a y extremum. This is done in sequence with new extrema being added on the right of the word. The code word for the letter in Figure 6 would be 10110010. The start position is labelled as a one so that the number of extrema is recorded, otherwise leading zeros would make the code word appear short.

In order to give information about where the extrema were detected another binary word, called the coord word (for coordinate word) is generated. The letter is divided into half in both the vertical and horizontal direction with the information obtained in the first trace. The four areas are labelled as shown in Figure 6. Whenever an extremum is detected it is appropriately recorded in the code word. In addition the label of the area in which the extremum was detected is recorded in the coord word. The coord word is formed in the same manner as the code word. That is, the two bits describing the coordinates of each extremum are added sequentially to the right side of the coord word. The coord word for the letter in Figure 6 is 0011110010101000.

In summary, the binary description of a character is very simply generated. It is based on a contour trace of the letter which is performed twice. The first trace determines the height-to-width ratio and the thresholds used in the smoothing operation. During the second trace the smoothing and the code word and coord word generation are done simultaneously. When the second trace is complete all pertinent information has been taken and the scan moves on to the next letter.

7.3 Recognition of the Unknown Character.

Recognition of the unknown character which is equivalent to classifying the character is done in three sequential steps with many comparisons at each step.

The code word obtained from the unknown character is compared with the list of code words of the previously learned characters. When the code word is found, the coord word of the unknown character is compared to all the coord words associated with that code word. When the coord word is found the height-to-width ratio is compared to the height-to-width ratio associated with the found coord word. When a match is found the character is said to be recognized and the appropriate character is given as output. If at any step a match is not found a signal is given signifying that the unknown is unrecognizable.

It is possible that an unknown may be recognized by its code word alone. That is, a certain code word may uniquely define a character. In this case the search can be terminated after the code word has been found without any reference to the coord word or the height-to-width ratio at all.

For the purposes of efficient data handling the maximum length of the three binary words were fixed. A maximum length of 12 bits for the code words was found experimentally to be sufficient. The maximum length of the coord word was fixed to be 18 bits instead of 24. It was found that if the letter was so complicated that it had more than 9 extrema the code word itself carried enough information to determine the letter uniquely. The height-to-width ratio is divided into four classes and thereby requires a two bit word.

The total information about the unknown letter is thereby contained in 32 bits of information. Most of the multifold machines in use today use a factor of 10 more information to describe the character in binary form immediately after scanning. Since in this technique a large part of the variability of the unknown is eliminated by the scanning technique, a great deal of data reduction has taken place in the scanner itself. Finally

the amount of information needed to describe the character will be reduced to about 6 bits. The ability of this technique to make a large amount of data reduction in the scanning equipment and a small amount in the processors is one of the main reasons why this technique can be made very inexpensively.

7.4 Significance of the Binary Description.

As mentioned in Chapter 4, the strokes or lines of a character define it. The extrema of a closed line are a minimal amount of information needed to describe it. In order to find the extrema, a stroke sequence for the character was defined by contour tracing the character. This is only one of many possible stroke sequences which can be generated from the printed character. It can be seen that the contour of the letter carries enough information to describe the letters by having humans draw the alphabet using this stroke sequence and reading the result. The sequence of extrema occurring during the trace of the character and the location in which they occur also carry enough information to describe the letter.

In Figure 7 the letter traced in Figure 6 is reconstructed using the exact position of the point where the extrema were detected. In this construction only

two rules need to be remembered. The first is that the trace line can never cross itself. The second is that in order to detect an extremum the trace must have receded from it exactly one threshold. Straight lines are used in the reconstruction, but smooth curves would be more letter like. The reader can verify that there is little freedom in the length and orientation of the lines used in the reconstruction. It is clear that enough information is retained to recognize the letter. The code word is significant, then, in that it carries all the information about the sequence of extrema.

It was found experimentally that the coordinates of the point where an extremum was detected do not need to be positioned accurately. It was found in fact that the coordinates need only be placed in one of the four areas described earlier. This is an experimental fact and not one that can be proved by reconstruction. Its significance can be shown, however, by consideration of the letters "D" and "P" which appear in Figure 8. The code words of both letters are the same. This portrays the fact that both letters have a smooth bulge on the right side. The coordinate words are different, however, and portray the fact that the bulge of the "P" is about one half the size of the "D" and occurs in the upper half

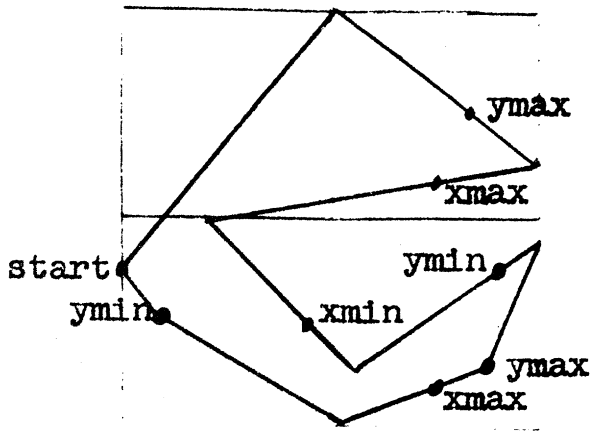


Figure 7.

Reconstruction of Unknown.

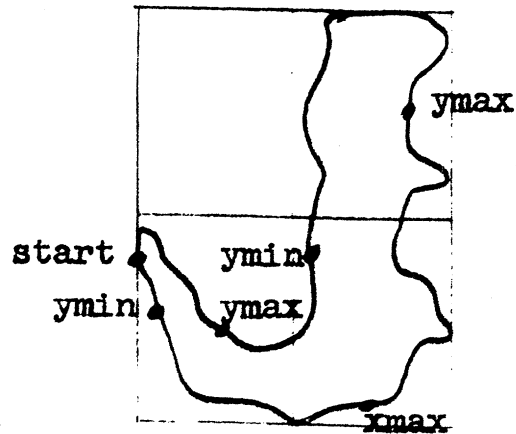
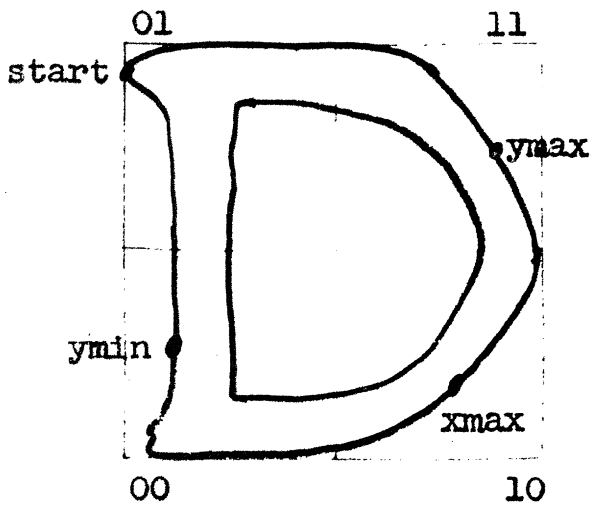


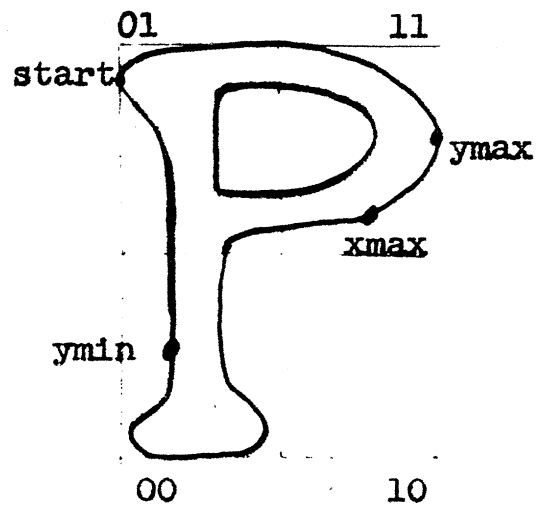
Figure 9.

Detected Extremum Position.



CODE WORD 1 0 1 0

COORD WORD 01 11 10 00



CODE WORD 1 0 1 0

COORD WORD 01 11 11 00

Figure 8.

Significance of Coord Word.

of the letter.

It is quite possible to refer the position of the detected minimum back to the original extremum. This, however, is not desirable as can be shown with a picture of a noisy "J" in Figure 9. If the x maximum were referred back to the extremum it could occur at a noise peak in either half of the letter causing confusion in the coord word. If the point is left where it is detected, however, it will always occur in the lower right area.

The height-to-width ratio is an obviously significant feature. Many times a human reader recognizes a letter merely by its height-to-width ratio and contextual information.

CHAPTER 8

Properties of this OCR Technique

8.1 Printing Noise Elimination

The hysteresis smoothing technique with the one-fourth character extent smoothing threshold is the basis for the success of this OCR technique. This operation virtually eliminates the major printing noises, that of style variations and ink run, in the scanner itself. Since the variability of the letter being scanned is greatly reduced by the scanner a very simple and inexpensive recognition device could be used.

A noise that is particularly bad for any contour tracing scheme is that of broken or touching characters. Fortunately this is a very minor noise, occurring only rarely. The occurrence of this noise can be even further reduced. If there is a lot of ink run on a certain page, characters will touch. In this case the threshold which determines black and white in the scanner should be set lower. If on the other hand, there was not enough ink while printing a page, there will be a lot of broken letters. These breaks can be filled in by defocusing the character and readjusting the black to white threshold. This will of course fill in the gaps between characters if the break in the

character is so large that a great amount of defocusing must be done.

The "salt and pepper" printing noise is also virtually eliminated since the scanning takes place only along the edge of the character.

The noises of rotated and displaced characters will be discussed in the next section.

8.2 Character Size, Orientation, and Position

The recognition is not based on the size of the character, the height-to-width ratio is a ratio and actual size is eliminated. In addition the threshold is set individually for each character and to a fraction of its extent. This allows complete freedom on the character size.

The binary description of the letter is relatively free of trouble due to rotation as well. A clearly printed character would have to be rotated so far that the top and bottom of a vertical stroke were displaced by one-fourth the character width in order to make an extraneous extremum possible. Even when rotated this far most characters would not introduce extraneous extrema. Assuming a height-to-width ratio of two, which is relatively high, this corresponds to a rotation of about 7 degrees. This is a very large

rotation. For example, if a line of type were printed with a tilt of the correct slope to cause this rotation of the letters, the tenth letter of the line would be displaced vertically the height of a letter, assuming normal letter spacing. It is possible that an extra extremum could be introduced with just a slight rotation, if the threshold was nearly reached in the untilted character. This case will be rare and can be handled by recording both code words.

It is possible that even with a slight rotation that the coord word changes. This would be due to the point of detection moving into another area of the character. However, even with the extreme rotation mentioned above, the maximum point shift could only be one-eighth the letter width. This is so small a distance that both coord words could possibly occur even without rotation due to the printing noise and both would have to be recorded. The major source of error due to rotation, then, is the introduction of extraneous extrema which requires rotation of approximately 7 degrees.

The recognition of a character is independent of its position.

8.3 Low Bit Rate.

The total information about the unknown character immediately after scanning is contained in 32 bits of information. This is about a factor of 10 smaller than that required by most multifont character recognition machines to date. This small amount of information is a witness to the fact that the threshold smoothing technique eliminates a large amount of the variability that can occur in an unknown. In addition it shows that this technique is a very efficient method for reducing printing noise. Finally, of course, the information will be contained in 6 bits so some machine operation is needed. This is done also in a very efficient manner using three table searches. Due to the small data rate and the efficient processing a very simple recognition machine can be used.

8.4 Speed

The speed of this recognition technique is quite sufficient to supply a reading machine. The major part of the time spent on a character is the time spent to trace the character twice. On the average two or three points must be examined at each point on the contour to find the next point in the trace. Let us assume that there are 100 points on the contour and three

points are examined each time. For two traces, then, 600 points are examined. A good flying spot scanner can be positioned and test a point in less than one microsecond. This implies less than 600 μ sec per character. This is about 1600 characters per second which is way above the speed needed for a reading machine and, in fact, in the normal speed range of optical character recognition equipment used in Electronic Data Processing.

8.5 No Feedback

The generation of the binary word for an unknown character does not depend on information from the recognition equipment. When the scan of one character is complete it moves on to the next character directly.

This is a desirable quality for a reading machine since, as was pointed out in Chapter 2, a sophisticated display can be time shared by many reading machines without losing the qualities of a personal machine.

8.6 Overlapping Characters Discriminated

Because of the nature of the captive scan, character pairs in which one character extends into the same columns occupied by the other without touching it can be separated and recognized. When the second trace of a character is complete, the column scan starts at

the bottom of the column one to the right of the right most column occupied by the character just scanned. If a point in this column is black, the tracing routine is started at this point. The point of this first trace which occupies the bottom point in the most left column of the trace is used as the starting point for the second trace. In this way letters are separated which do not have a completely white column separating them.

8.7 Low Cost

In order to compare the cost of this character recognition technique with that needed in a reading machine a rough estimate was made on the parts alone cost. This estimate includes: an opaque scanner with enough resolution to read the print of common books, papers, and magazines, including the necessary optical equipment; the electronic logic needed to direct the trace and calculate the height to width ratio, the code word, and the coord word; and a magnetic drum for use in the table search for recognition, including the input and output electronics. In other words, this estimate is intended to cover all that is necessary to translate the printed page into a sequence of voltages of the recognized letters. The parts cost of this equipment was estimated to be about \$2,500.

This amount is within the goals set for that of a reading machine, about the price of a car. The necessary parts needed to get spoken letter output are minimal since a magnetic drum already exists and the voiced letters could be recorded on it either digitally or analogue. More complicated outputs would cause the price to increase, that is up to the discretion of the builder.

8.8 Simple Tactile Display.

If a letter is not recognized by the machine a simple tactile display must be generated. One possibility is to display the trace of the letter tactilely as it is generated. This would give the outline of the unrecognized character for human recognition. Another possibility is to display the points where the extrema were detected and connect them with straight lines. This would be equivalent to that of Figure 7 in Chapter 7. It is felt that the coordinates of these points must be displayed with more accuracy than that required in the coord word. This must be decided by experimentation. Both of these possibilities are simple displays based on the data picked up from the character. No extra data need be taken for human display.

CHAPTER 9

Experimental Data

9.1 Purpose of the Data.

Experiments were run to test the validity of the character recognition technique, described earlier, for use in a reading machine. It was the purpose of these experiments to show that this character recognition technique was valuable as the major technique on which to build a character recognition system. The data is intended to show that this character recognition technique in itself, although not perfect, meets the requirements of the reading machine as described in Chapter 3. In addition it will be shown that improvements can increase the capabilities of the equipment well beyond that needed for a reading machine. None of the improvements which require a change in the basic technique as described are incorporated in the data. This was done purposely so that the technique can be tested on its own merits. The possible improvements will be described in Chapter 10.

9.2 The Experimental Setup.

Unfortunately an opaque scanner was not available for experimentation. Instead a scanner was used which requires a transparency. A transparency of the print to be recognized was taken using Polaline 146-L film. This

is high contrast film, but this fact is unimportant since a threshold had to be set in the scanning setup to make a decision between black and white. The time of exposure sets the threshold of the film. The cathode-ray tube used was that of the TX-0 computer. A spot was positioned on the screen and then intensified. This spot was focused onto the transparency and the light transmitted through the transparency was columnated by a lens system and measured by a photo-multiplier tube . The current of the photo-multiplier tube was then tested against a threshold to determine whether the point was black or white.

In the experiments performed, a line of print on the picture was selected by eye, the computer then scanned this line of print row by row and stored it in core storage.

The trace is performed by testing the locations in core rather than the picture of the print itself. This is certainly not a requirement, but it was felt that for the experimentation the quality of the scanning equipment should be divorced as much as possible from the recognition technique.

9.3 The Learning Program.

Two separate main programs were written. One program compiled the data on the characters scanned. The other

program used the data previously learned to recognize unknown characters and print them out.

The learning program performed the following operation. A column scan was generated which moved from left to right until the first black area was found. This black area was then contour traced for the first time, and the program recorded all points in the trace in display format. When the first trace was completed the complete trace was displayed on the cathode-ray tube and the typewriter typed a message asking what the character encircled is called. When the appropriate character was typed, the computer traced the character the second time computing the code and coord words. Then the code and coord words and the height-to-width ratio were recorded in the location in core associated with this character and they were typed out on the typewriter. In each case if the code word was not encountered before for that character an appropriate message was typed out. The scan then moved on to the next black area and the sequence was repeated. If the black area in question was not a character of interest an appropriate key was struck and the scan moved on to the next black area.

When sufficient data on the characters was recorded a program was written to recognize unknown characters on

the basis of this data.

The program operates as the description of the recognition procedure describes. The operation could be set in two modes. In the first mode the typewriter types out the code word, the coord word, and the height-to-width ratio. Then it types out the character to which the data corresponds. If the data corresponds to no previously learned character the unknown is called a glob. In the second mode the typewriter only types out the sequence of characters. If a black area is unrecognized an underline is typed. The program also counts vertical blank columns and if it exceeds a certain number, which is set manually, a space is typed.

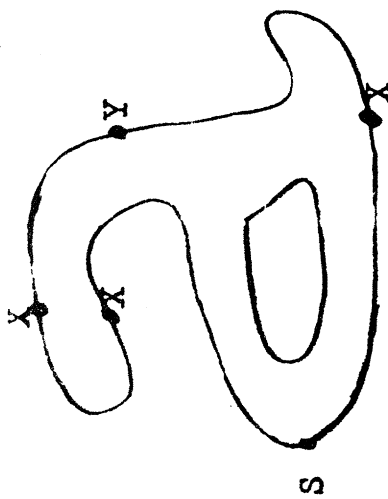
There is no reason why both these programs could not be made into one. This could be incorporated very well into a reading machine. If a character is unrecognized it could be displayed tactilely and the reader could recognize it and inform the machine so that the machine would recognize it the next time it was encountered. This would be of great value in reading material where certain uncommon symbols appeared often.

9.4 Multiplicity of Code Words.

Due to the particular way in which code words are generated it is not expected that each character will

have only one code word. If for example the trace decreases from an extremum along a line of 45 degrees elevation it is quite possible to have a y extremum and an x extremum detected at approximately the same position. In this case either extrema may occur first along the trace due to the particular noise. In this case both code words will have to be stored. As an example of this phenomenon, Figure 10 shows four different code words associated with the letter "a" found in the New York Times. The four code words occur because of two separate variations. The start position may occur closer to the bottom of the letter than the threshold. In this case no y minimum will occur at the end of the trace. If the start position is higher than one threshold on the letter, a y minimum will occur at the end of the trace. The other cause of a change in the code word is that in some cases the hook at the top of the "a" is longer than one threshold and sometimes it isn't. When the amount of hook exceeds the threshold a pair of y extrema are introduced. The combination of these two possible variations causes the four code words shown in Figure 10.

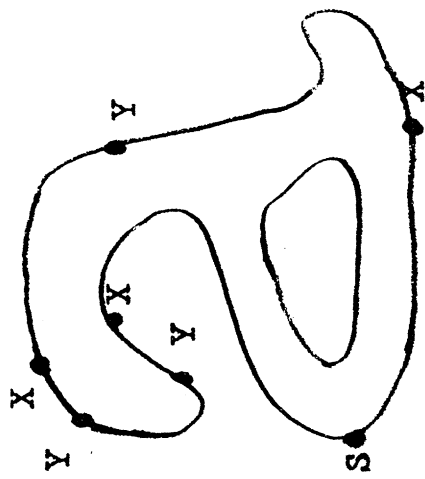
In much the same way additional coord words are formed. Whenever an extremum falls near a character



CODE WORD 11101

COORD WORD 000101110

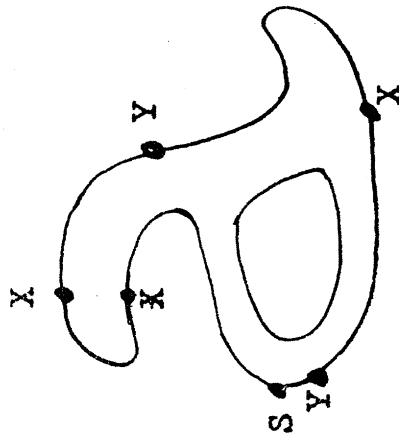
FOUND 25 TIMES.



CODE WORD 1100101

COORD WORD 00010101110

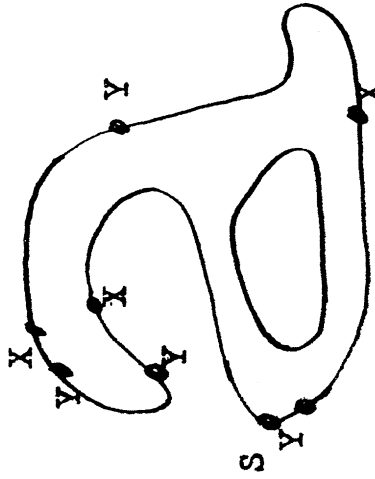
FOUND 16 TIMES.



CODE WORD 111010

COORD WORD 00010111000

FOUND 7 TIMES.



CODE WORD 11001010

COORD WORD 0001010111000

FOUND 7 TIMES.

Figure 10. Letter "a" Found in the New York Times.

division line, noise will force two coord words corresponding to the extremum position being on either side.

9.5 Confusion Among Letters.

Because the recognition of an unknown character is based entirely on the outside contour, confusions can result when the characters are distinguished by an interior feature. Fortunately this is very seldom the case in the Roman alphabet. One notable example of this is the distinguishing feature of the capital "B" and "D". If the slight dip at the right edge of the "B" is smaller than one-fourth the letter width, and it often is, the two letters will be identically classified. These two letters with their detected extrema are shown in Figure 11a. In addition it is quite possible the capital "O" will be confused with these letters due to printing noise. In Figure 11b is shown the letter "O" with the printing noise added which confuses it with the letters "B" and "D". The amount of printing noise needed to confuse these letters is relatively small.

Later the confusions between letters obtained from actual print will be listed in detail. The number of confusions obtained is a surprisingly small number. In every case it is possible to distinguish these letters with a relatively small amount of extra operations on

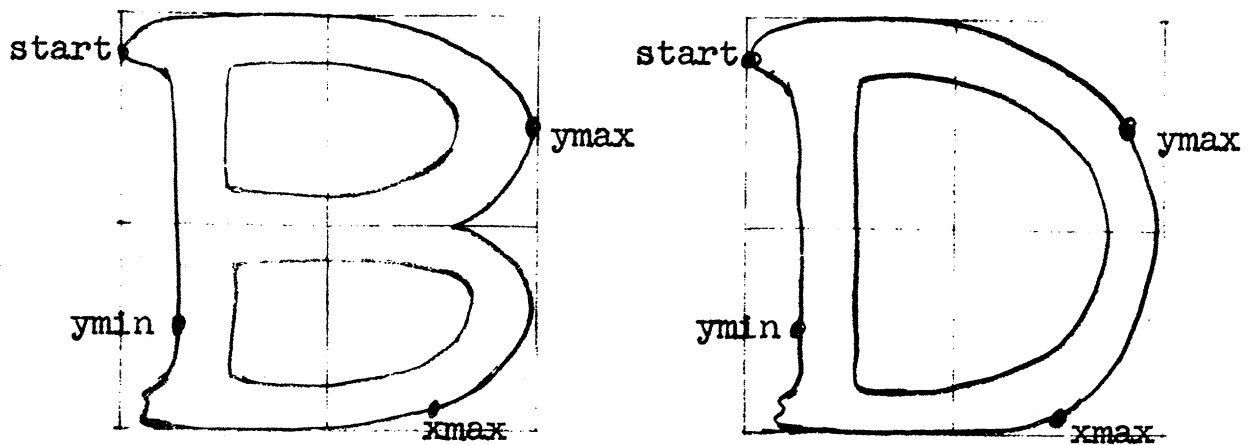


Figure 11a.

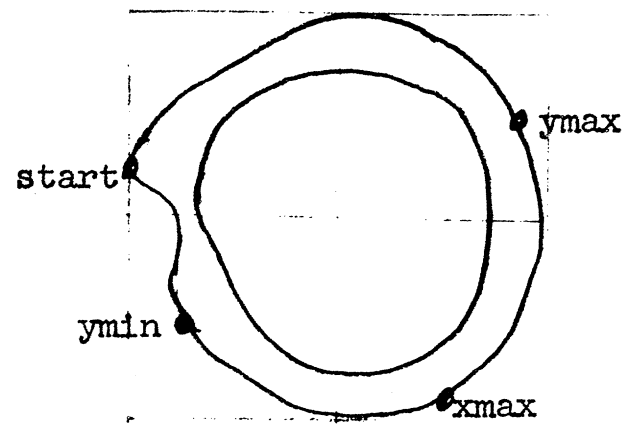


Figure 11b.

Particular Characters Which Require Additional Tests For Separation.

Figure 11. Confusion Among Letters.

the unknown. Fortunately most of these confusions occur with capital letters which are much less frequent than lower case letters. For this reason the reading rate is well within the limits specified even with the confusions which occur.

9.6 Tests on Standard Alphabets.

In order to make a realistic test on the variable font capability of this character recognition technique, a set of ten different popular alphabets were compiled. By correspondence with the Mergenthaler Linotype Company and the Lanston Monotype Company a list of the most popular type fonts in use today in the areas of books, magazines, and newspapers was compiled. From this list ten type fonts were chosen which represented the three areas. These type fonts are Baskerville, Caladonia, Corona, Excelsior, Granjon, Ionic, Janson, Opticon, Primer, and Times Roman. These type fonts are all of the Roman style but are quite different in design.

Transparencies of these type fonts were taken from actual printed material so printing noise is included.

These ten alphabets were used also to experiment with various threshold settings. For each setting of the thresholds the alphabets were scanned and the results compiled together. Based on this data the letters were

then classified in terms of their code words, coord words, and height-to-width ratio. The number of code words encountered was recorded as well as the number of coord words needed to distinguish letters which had the same code word. This is a measure of the amount of storage required to recognize the letters. The number of confusions represents the number of letters which could not be distinguished. If 10 "B's" fell into the same classification used for "D's", this was called 10 confusions.

These statistics are listed in Figure 12 for comparison. In order to make the results meaningful it was assumed that ascenders and descenders of the letters were detected. This would certainly be done in any reading machine using this technique. The method for obtaining these is straight forward (see Earnest⁽¹⁹⁾). It can be done even when there are only several letters in a line of type. Although almost all letters are distinguished by their contours alone the i, j, and l of the mixed type fonts are more readily distinguished with ascender and descender detection. Rarely other letter pairs are distinguished as well.

In Figure 12 three separate confusions are calculated. The combined confusions are the confusions obtained from the 520 upper and lower case letters of the 10 type fonts. .

The upper case confusions are those obtained from the 260 upper case letters. The lower case confusions are those obtained from the 260 lower case letters. If the last two numbers do not sum to the combined confusions this implies that a lower case letter was confused with an upper case letter. The error rate for lower case letters is the error rate that would result if all lower case letters were equally likely to occur. This number is given since by far the majority of letters in common text are lower case letters. This number is not the error in reading rate, however, since the probability of the occurrence of the letters in common prose must be included. This will be demonstrated with an example in the next section. It is very significant that the confusions are not random substitutions of the 25 other letters of the alphabet, but with one other letter. This implies that the confusions can be easily coped with.

The first run was made with the thresholds set individually to one-fourth the letter extent in each respective direction. The second run was identical to the first but with the thresholds set to three-sixteenths the letter extent. This threshold was small enough to detect many of the serifs. For that reason there were more code words and some were longer than 12 bits. The number of combined

	CODE WORDS	COORD WORDS NEEDED	COMBINED CONFUSIONS	UPPERCASE CONFUSIONS	LOWERCASE CONFUSIONS	ERROR RATES FOR LOWERCASE
1. Tx = 1/4 l.w. Ty = 1/4 l.h.	60	73	39	39	0	0.0 %
2. Tx = 3/16 l.w. Ty = 3/16 l.h.	110	70	17	13	3	1.15%
3. Tx = 1/4 l.h. Ty = 1/4 l.h.	67	69	52	49	3	1.15%
4. Tx = 11/32 l.h. Ty = 11/32 l.h.	45	62	43	37	0	0.0 %
5. Tx = 3/16 c.h. Ty = 3/16 c.h.	115	75	10	9	1	0.39%
6. Tx = 1/4 l.w. Ty = 1/4 l.h. V.D. = 9/32	61	67	31	30	1	0.39%

Tx = Horizontal Waveform Threshold; Ty = Vertical Waveform Threshold
 l.w. = letter width; l.h. = letter height
 V.D. = Vertical Division; c.h. = capital height

520 letters scanned in each run.

Tests made on: Code Word; Coord Word; Height-to-Width Ratio;
 plus presence of Ascenders and Descenders.

Figure 12. Letters From 10 Different Alphabets.

confusions reduced however. In the third run both thresholds were set to one-fourth the letter height. Although the number of code words needed is approximately the same as the first run the number of confusions is higher. This occurs because some significant horizontal features in narrow tall letters such as the cross bar in the lower case "f", are smaller than serifs and are eliminated. The fourth run is the same as the third run but with a larger threshold. The fifth run is quite different in that the thresholds are set once for all the letters. In this case both thresholds were set to three-sixteenths the height of an upper case letter. If the thresholds were increased to one-fourth the results would be close to those of run four. The sixth run is the same as the first run except that the letter is not divided in half vertically to generate the coord words. It was noticed that in some cases, notably in the letter "P", the bulge of the letter, although shorter than the height of the letter, was larger than one half the letter and was not detected as being short. With the division set at nine-thirty seconds the letter height from the base, all these short bulges were detected.

Considering the fact that the larger the threshold the more noise is rejected, possibly the settings used

in the sixth run are best. In a reading machine for the blind the threshold will be adjustable so that it can be trimmed up to fit the type style. Any one of the settings in Figure 12 would produce an acceptable reading rate for a reading machine.

As an example of the type of confusions that occur, those of the sixth run will be listed. In this run the following confusions were made: B for D, Q for O, R for A, Y for V, c for e, H for N and C for G. It is interesting to note that the machine confuses letters which are readily confused by eye. Possibly the two most surprising confusions were H for N where noise made the cross bar appear slanted and C for G. The cross bar on the G is always more than one-fourth the horizontal extent of the letter, however, the line leading to the cross bar is in general very thick and hides the fact from the contour trace. All of these letters can be distinguished by further operations which will be discussed in Chapter 10.

9.7 Example of Magazine Reading.

In order to test the validity of this character recognition technique for reading machine applications a full page of Time Magazine was read. To be as impartial as possible the first page of text was taken from the

Time Magazine of June 11, 1965. Over 3300 characters were encountered. There were no touching letters and only one broken letter. The page was scanned once and the data from all the characters was compiled. From this data the appropriate division for the four classes of height-to-width ratio was set and the characters were classified according to these code words, coord words, and height-to-width ratio. The number of points around the contour of a letter was counted. If it was less than 64 points it was called a general punctuation mark and no further data was recorded. The page was then read from the data collected. This is equivalent to reading with a completely trained machine.

Figure 13 shows the confusions that occurred. The titles of the sections were printed in Sans-Serif type and are kept separate. The confusions are separated into three sections: letters, punctuation, and numbers. The letters had first preference in classifications and the numbers last. Number confusions with letters cannot be considered errors as stated in Chapter 3, since a number can be sensed from context easily. It is surprising, however, that the number 0 is confused with the letter "d" instead of the letter O. This is a peculiarity of the particular type font. The number 0 was very narrow so

that its height-to-width ratio was confused with "d". The only other surprising substitution is when the dot above an "i" was broadened by noise and was detected as a punctuation mark. This happened three times.

Unfortunately the property of detecting ascenders and descenders did not exist in this reading program. Therefore the reading error rate is higher than it need be. The error rate including letters and punctuation from the titles and text is about 3%. If ascenders and descenders were detected the reading error rate would drop to about 1%. Figure 14 is the text as recognized with this machine. It is very easy to read if the possible substitutions are kept in mind, as the reader can verify. This is due to the fact that the substitutions occur between only two or three letters instead of all 26. All punctuation marks which are small are replaced by the low asterisk independent of the vertical height at which they were detected.

9.8 Reading After Short Training.

In order to demonstrate the ability to read after a short training passage, the last one-third of the same page of Time Magazine was read after training on the first two-thirds of the page. The results of this reading are presented in Figure 15. Essentially the same confusions

Letters

Text	Titles (Sans-Serif)
30 i → l	1 r → p
2 l → i	3 a → o
6 j → l	2 i → l
16 r → p	1 c → e
1 p → r	
24 c → e	
1 C → e	
2 B → D	
1 B → d	
1 a → D (broken a)	

Punctuation

1 space → .	2 space → .
2) → l	

Numbers

9 0 → d
9 1 → i
3 1 → l
1 4 → q
4 5 → S

Data From Over 3300 Characters.

Figure 13. Confusions From Time Magazine, June 11, 1965.

SpACE

CloSl*ng the Gop
(See Cover)

He stood on top of his spaceship*s
white titanium hull*. He touched it with
his bulky thermal gloves*. He burned
around like Duck Rogers propelling
himself with his hand*held jet*. He
floated lazily on his back*. He loked and
laughed*. He gazed down at the earth
id3 miles below*. spotted the Houston*
Galveston Day area where he lives and
tried to take a picture of it*. Like a gas
station Dttendant* he echecked the space*
craft*s thrusters* wiped its wndshield*
Ordered to get back into the capsule* he
protested like a scolded kid*. **I*m doing
great*** he said* **It*s fun* I*m not eoming
in*** When* after 2d minutes of space
gymnasties* U*S* Astponaut Edward
Higgins White II* 34* finally did agree
to squeeze himself baek into his Gemini
q ship* he still had not had enough of
space walking*. Said he to Command
Pilot James Alton McDivitt* **It*s the
saddest day of my life***

White*s exhilarating space stroll pro*
vided the moments of highest drama
during Gemini 4* scheduled 62* orbit*
98*hour* i*7dd*ddd mile flight*. White
spent twice the time outside the space*
eraft that Soviet Cosmonaut Aleksei
Leonov did last March 18* and he had
much more maneuverablllly* all Leonov
did was somersault around at the end of
a tethep* getting dizzy* while White
moved around pretty much at will*.

Seend Genepotl*on*. Still* Gemini*s
planners would have serubbed White*s
EVA (for Extra*Vehicular Actlvityl
expedltion in a second if they had
thought it might detract from the
flight*s basic missions*.

Figure 14. Example of Reading Time Magazine

In Gemini 4* the U*S* took a big step toward closing the gap in the man* in*space race* in which the Soviet Union got off to a head start*. More important* the flight signaled the advent of the second generation of U*S* spacecraft and spacemen*. The two*man Gemini capsule is to the old Mercury capsule what a thunderbird is to a Model T*. Al* most all previous U*S* space flights were pre*planned to the second* and any de* viation meant trouble*. In Gemini 4* the astronauts were given considerable flex* ibility* could and did change their plans and improvise at short notl*ce*. For the first time* a U*S* space flight was con* trolled from Houston*s supersophisticat* ed Manned Space Center* which makes Cape Kennedy almost as obsolete as a place once called Canaveral*.

Moreover* the space men themselves were second generation*. Project Mer* cury*s pioneers were national legends almost before they got off the ground*. Yet who* before last week* knew very much about Jim McDivitt and Ed White?

The Team*. The pair made an almost perfect space team*. Inside Man Mc* Divitt is a superb pilot and a first*class engineer who is the son of an electrical engineer*. Outside Man White is a daring flyer*a fine athlete* a military career* man who is the son of a retired Air Force major general who flew every* thing from balloons to jets*.

McDivitt* whose 36th birthday is this week* is a whippet lean (5 ft* 11 in** is 5 lb*) Air Force major*. As a youth* he did not seem exactly the type to be a space ship looney*. After graduating from high school in Kalamazoo* Michigan** he worked for a year as a furnace re*

palrman* then drifted rather aimlessiy
 into tlny (then S3i studentsl Jackson
 Junior College ln 1948*. On his college
 application he wrote* **I think I would
 like to be an exrlorer and a novelist***
 A so*so student* McDivitt finished his
 two*year eopause in 195d* and since he
 was about to be drafted into the Army*
 declded he might as well loin the Air
 Foree as an air cadet*. He found a
 home and a ealling*.

As a jet fighter pilot* he went to Ko*
 rea* flew i45 combat missions* won
 three Distinguished Flying erosses and
 five Air Medals*. In 1957 the Air Force
 sent hlm to the University of Michigan
 to get a degree in aeponautical engineer*
 ing*. dy now more mature and sure of
 himself* he got stralght A*s* graduated
 first ln an Engineerlng School class of
 6d7*. From Miehigan he went to the
 Experimental Test Pilot School at Ed*
 wapds Alr Force Base in Californla*. was
 selected fop the X*is testing pprogram*
 but applied instead for Gemini*.

He was plcked wlth elght others_
 i*nclding Ed White in September
 1962*. Jim McDivitt sounds about as dis*
 passionate about being an astronaut as
 he would about fixing furnaees* **There*s
 no magnet drawing me to the stars*** he
 says flatly* **I look on this whole prolect

Figure 14. Continued

exist as in the previous example. The major difference is that some characters are not classified at all. This means that the binary data obtained from these characters on the last third of the page did not appear at all in the first two-thirds of the page. In almost all cases this occurs because the training period was not long enough. The unclassified binary descriptions would occur often enough in a long sample to be significant. This is evidenced by the fact that most unclassifications occurred for upper case letters and numerals. In a few instances, as with one each of the lower case "a" and "r", the particular unclassified binary description was caused by a certain noise pattern and possibly would not occur again, even in a very long sample. These are the cases where even a trained machine would not have learned them and they would have to be presented, if needed, in some alternative display mode.

The data portrayed in Figure 15 points out the obvious fact that, after training termination, binary descriptions occurring for the first time are not classified.

In a reading machine for the blind the characters which were unclassified would be presented in some other mode. If the reader recognized the character he would inform the machine and a classification of the binary

description would be formed. In this way the recognition rate would improve as more and more print was read.

Figure 15 shows how well the machine can recognize after reading 2200 characters of text.

Some characters appear so seldom that a lot of text would have to be read to learn all the variations of the binary description. Possibly many typical variations could be learned from one appearance of a character if certain moderate distortions were performed on the character the first time it was encountered, in order to generate more than one binary coding upon first encounter. One operation might be a slight rotation of the character. The four different classifications of the letter "a" appearing in Figure 10 could possibly be obtained from one appearance of an "a" if it were rotated slightly to a few different orientations. A few well chosen operations applied to the characters while learning might speed up the learning process considerably.

9.9 Validity of this Technique for Reading Machine

Applications.

The requirements for OCR equipment for use in a reading machine were listed in Chapter 3. The properties of the technique under discussion will be compared to these requirements.

McDivitt, whose 3rd birthday is this week, is a whippet lean (5 ft. 11 in., 150 lb.) Air Force major. As a youth, he did not seem exactly the type to be a spaceship looney. After graduating from high school in Kalamazoo, Michigan, he worked for a year as a furnace repairman, then drifted rather aimlessly in Ontario (then State) Jackson Junior College in 1944. On his college application he wrote: "I think I would like to be an explorer and a novelist." A so-so student, McDivitt finished his two-year course in 1945 and since he was about to be drafted into the Army, decided he might as well join the Air Force as an air cadet.

As a jet fighter pilot, he went to Korea, flew 14 combat missions, won three Distinguished Flying Crosses and five Air Medals. In 1947 the Air Force sent him to the University of Michigan to get a degree in aeronautical engineering. By now more mature and sure of himself, he got straight A's, graduated first in an Engineering School class of 1947. From Michigan he went to the Experimental Test Pilot School at Edwards Air Force Base in California, was selected for the X-15 testing program, but applied instead for Gemini.

He was picked with eight others, including Ed White in September 1952. Jim McDivitt sounds about as dispassionate about being an astronaut as he would about fixing furnaces. "There's no magnet drawing me to the stars," he says flatly. "I look on this whole project

Figure 15.

Reading After Short Training

The speed requirement is very easily met. The recognition can be done a factor of 10 or even 50 faster than required for a reading machine.

The requirement on error rates is that there be less than 10% substitution error rate. The example of machine reading of Time Magazine shows that the reading error rate is only 3% without ascender and descender detection and drops to 1% with this facility. In addition these errors are not strictly substitution errors since only one or at most two letters are confused. That is, if a letter is in error the correct letter is one of two other letters not one of twenty-six.

The numerals are not confused among themselves but are confused with some letters which are topologically similar. This is quite satisfactory according to the requirements.

The data from the tests of the ten standard alphabets in popular use demonstrates the validity of this technique for use with varied fonts. At most a small trimming adjustment to cope with type style will be needed. The technique satisfies the error rate requirement with almost any reasonable setting of the thresholds.

The complete recognition equipments including the scanning equipment is estimated to be only a few thousand dollars meeting the requirement imposed by a personal reading machine.

CHAPTER 10

Areas for Further Research

10.1 Introduction.

In some specific reading applications the error rates obtained with the basic technique described earlier might be too high. A list of suggestions which would increase the recognition power of the basic technique is given in this chapter. Possibly one or two of these suggestions would lower the error rate considerably. The choice of which suggestion to use, if any, must be decided in part by the ease in which it is incorporated into the machine that is built.

10.2 Test for Ascenders and Descenders.

Testing for ascenders and descenders can readily improve the recognition rate as was demonstrated in Chapter 9. The letters l, i, and j look very much alike in some fonts and are distinguished by the ascenders and descenders. Many simple ways for detecting these features could be incorporated. Earnest⁽¹⁹⁾ has shown that the ascenders and descenders can be detected by essentially plotting the density of black in a line of type as a function of the vertical position. Two distinct peaks appear corresponding to the base of the line of type and the top of the lower case letters. Once these two lines

have been determined a descender or ascender can be detected during the trace by testing the vertical coordinates of the trace points against the coordinates of these lines. A buffer zone will have to be allowed, of course, to account for vertical misregistration.

10.3 Eliminate Start Position.

The start position of the contour trace and code generation as described is distinct from the other code points. The start position assumes a state of x minimum and y minimum. It will always be at an x minimum but may be at a y maximum as in tracing a "V". In addition the point occurs at the x minimum and not one threshold away as the other points do.

One way to eliminate this bugaboo is to always start the trace at a point where an x extremum is detected. Starting at an x extremum would start the code word with a one and thus label its length. This can be done by starting the second trace of the letter as usual and detecting the extrema without recording them until an x extrema is found. The start could now be assigned this point and the trace continued recording the code and coord words until this new start position is reached again.

Figure 16a shows how A and R would now be discriminated by the coord word even though they would not have been with

the start position as described earlier.

10.4 Test on Contour Size.

If the size of the printing can be normalized, the number of points on the contour might be a distinguishing feature. This has been used as described earlier, to detect punctuation marks. This would be of most use in discriminating upper and lower case letters.

10.5 Change of Areas for Coordinate Definition.

It was found in the experimentation that the recognition rate increased if the vertical division did not occur at one-half the character height. Possibly a change in the horizontal division would be effective as well. In some cases it might be advantageous to divide the character into non-symmetric areas.

10.6 Test for Leading Vertical.

It can always be assumed that the second trace start at the bottom rather than the top of a leading vertical line. This can be done by effectively tilting the character. Assume that the first trace of a character begins at the point x_s, y_s . The x coordinate of every succeeding point is tested to see if it is less than $x_s - k(y_s - y)$ where y is the y coordinate of the tested point. If x is less than this value it replaces x_s . The final x_s and its corresponding y coordinate.

determines the starting point of the second trace. The factor k determines the effective tilt of the character. In general it would be very small.

If it is assumed that the starting point is at the bottom of the leading vertical of any character that has one, the vertical can be simply detected on the second trace by testing to see whether the y coordinate of the trace point reaches nearly the height of the letter before x increases one threshold unit. If it does a leading vertical must exist.

This test would distinguish D from O and R from A, as well as some other character pairs.

10.7 Test of Character Interior.

There are several ways to test the interior of a character. Two ways will be mentioned here.

An important property of the hexagon tracing described earlier is that edge extraction can be done. If the character is scanned and stored it is a simple matter to erase each point as it is encountered in the second trace. This leaves only the interior edges plus perhaps a few scattered points. It is not necessary to store the character, however, since every point that has six black nearest neighbors is set to white this test can be made during the trace. It is impossible to erase points this way, however, and the inner edge would have to be found another way.

A much simpler way to test the interior of a character can be made by a single vertical sweep through the center of the character. The extent of the character in horizontal direction is known after the first trace so the center of the character in the horizontal direction can be found. At this point a vertical sweep can be made through the letter and the number of black and white crossings counted. "Salt and pepper" noise would spoil this measurement. As pointed out earlier, "salt and pepper" noise is almost non-existent and would not cause serious difficulty.

This test would distinguish B from D, c from e and some other characters which look much alike when there is a lot of ink run.

10.8 Relative Positions of Neighboring Extrema.

A possible test which may be very difficult to design is that of relative positions of neighboring extrema. This is best described by the examples of the noisy c and e in Figure 16b. Even though the code and coord words are identical for these two letters the x extrema is much lower than the y extrema in the upper right portion of the e. This is not the case for the c.

10.9 Tracing in Reverse Direction.

In some cases characters can be distinguished by tracing in a counter clockwise direction. If this is done with an

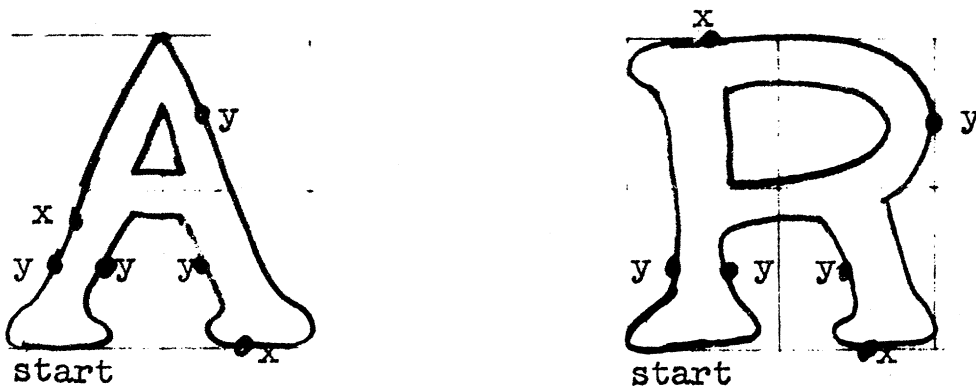


Figure 16a. Start Elimination.

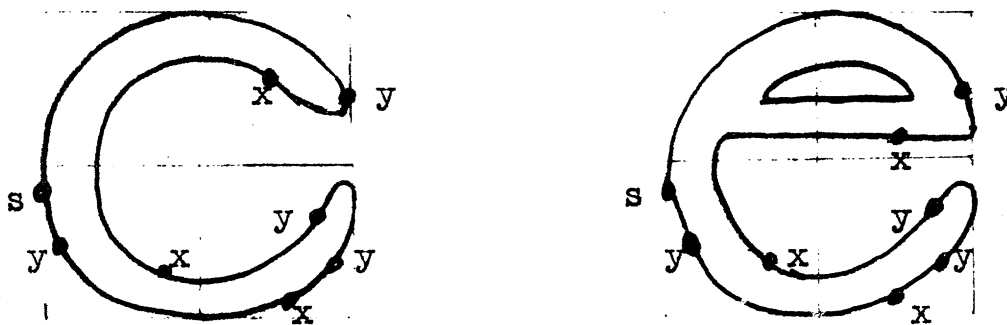


Figure 16b. Relative Position of Extrema.

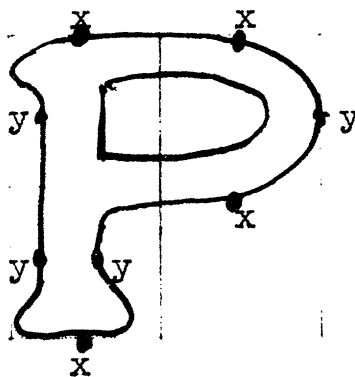


Figure 16c. Extrema Pairs Generation.

Figure 16. Example of Technique Variations.

upper case G an x extremum will be detected on the cross bar itself. The reader can verify that this will distinguish it from the upper case C if the relative positions of neighboring extrema are used.

10.10 Extrema Pairs Generation.

One possible variation of the tracing procedure would be to enclose the actual extremum between two points on the trace. The trace proceeds as usual but when an extremum is detected the trace is reversed until this extremum is passed in the other direction one threshold unit and this point is recorded as well. An example of a P scanned in this manner is shown in Figure 16c. There are two obvious advantages. Any extremity, such as the bulge on the right side, is bounded by two points. The position of the actual extremum can be estimated by adding one threshold unit to the x position where the two points detected the x maximum (subtract for an x minimum) to find the x coordinate and take the average of the y coordinates of these two points for the extremum's y coordinate. In this way the coord word would be much more stable. For example in the P the x maximum would always occur well in the upper half. Another advantage is that this procedure allows the measurement of the extent of the character feature on which the extremum occurs. This would help distinguish some characters,

notably the upper case V and Y where the two points detecting the lowest y minimum would be farther apart for the V than they would be for the Y.

A code word can be generated for this procedure in binary form just like the earlier tracing procedure. This procedure might raise some unseen problems, however, since it has not been tested.

10.11 Non-Orthogonal Directions.

The two directions, vertical and horizontal, were chosen because they seemed natural. There is no reason to believe they are the best choice of directions. It might well be that some other choice of directions might give better results.

10.12 First Prototype.

In some reading machine applications the error rate obtained with just the code word, coord word, and height-to-width ratio description might be too high. In this case one or more of the previous suggestions could be employed to increase the recognition rates. The type of improvement chosen will depend, in part, upon the ease of implementation into the previously existing equipment. For this reason the first prototype should be designed so that implementation of improvements could be accomplished easily.

On the basis of the experiments reported in this thesis, possibly the first prototype should be built as described herein. All possible variations such as threshold settings should be controllable by the operator. This is the trimming adjustment mentioned in Chapter 2. It is clear from the data that the threshold is not a critical adjustment and need not be changed drastically when a new type font is encountered.

It is possible that another tracing routine might be used in place of the hexagon tracing routine. It is important that the tracing routine be able to move from any configuration to the next trace point with the same simple operation, allows edge extraction, and that it always returns to the starting point after encircling the character once.

When adding improvements or making changes to the technique, it is important to keep in mind that the line defines the character and the line should thereby be defined by the technique. The technique should have this capability in the presence of the major printing noises, that of style variations and ink run. The technique described in this thesis was designed to eliminate these noises inexpensively and this property should not be destroyed. The major properties are listed in Chapter 8 and care should be taken that they are not compromised when changes are made.

CHAPTER 11

Summary and Conclusions

This thesis was concerned with the application of optical character recognition (OCR) to a very special problem, that of a personal reading machine for the blind. Since the proposed machine was, in fact, a personal one, interplay between the machine and the reader could be used to advantage. The requirements of OCR in this application were found to be quite different than the requirements in Electronic Data Processing. The most significant changes were the reduction of speed and error rate requirements and the increase in the requirements of low cost and the ability to recognize type fonts of various styles. No existing OCR techniques were found suitable for this purpose and a new technique was developed.

The technique used here is based upon tracing the outside black-white boundary of the printed character. It was found that in the Roman alphabet the outside contour contained enough information to describe the characters. The outside contour of the character is traced by means of a digitally controlled flying spot scanner. A column scan is generated which proceeds from left to right until a character is contacted. The scanner then traces the contour of this letter by means of simply generated

local operations until the trace is complete. The basis of description of this trace is the local maxima and minima of the horizontal and vertical waveforms of the trace as it proceeds around the character and the positions at which these extrema occur.

A method for smoothing had to be developed which would eliminate extraneous extrema due to printing noise, and yet preserve the extrema due to significant features of the character. The smoothing method used is a non-linear technique employing hysteresis. This technique eliminates many problems associated with printing noise. In effect, the smoothing technique picks out only features which extend at least one-fourth the character extent and completely eliminates all others. With this large noise threshold most of the variability of the characters due to style variations and printing noise is eliminated in the scanner itself. The gross features retained are sufficient for character recognition.

With such smoothing employed, then, the character is described by two consecutive contour traces. The first trace measures the horizontal and vertical extent of the character. With this information the thresholds are set independently for the horizontal and vertical directions and the height-to-width ratio is calculated.

The second trace then begins and the extrema are detected. Two kinds of data are extracted. The code word contains information about the sequence of extrema, and the coordinate (coord) word states in which of four areas of the character the extrema were detected. The total description of the character requires 32 bits of information, 12 bits for the code word, 18 bits for the coord word, and 2 bits for the height-to-width ratio.

The classification of the unknown character is done in three steps. The code word of the unknown is compared to the list of previously found code words. When a match is found, the coord word of the unknown is compared to all coord words that previously occurred with that code word. Finally the height-to-width ratio is compared and the appropriate character is given as output. The search can be terminated at any step if the unknown is uniquely classified at that step. If no match is found an appropriate signal is given. The data presented shows that the ten most common type fonts in use today can be classified all together with less than 100 code words and less than 100 coord words.

It is expected that the personal reading machine will operate in two or more modes. When the character is recognized the reading machine will operate in its

normal mode and present the character or the word in which it occurs to the reader. When the character is not recognized the reader will have the choice to go on to the next character or to be presented some recoding of the optical signals which were obtained. The reader can then recognize the character and train the reading machine if desired.

The blind reader will have complete control over the reading machine. If errors begin to occur regularly he will be able to adjust the position and controls of the optical pickup equipment, such as focus and threshold. In addition he will be able to adjust the scanning and recognition procedure. One such adjustment would be changing the smoothing threshold which can be varied from the one-fourth character extent to suit the print quality. Other changes could come from use of the procedure changes given in Chapter 10.

A reading machine such as this would give the blind reader a very versatile machine which he could tune in to suit the quality of the print being read. In such a versatile machine it is difficult to predict apriori how high a reading rate can be obtained. It is not expected that the reading machine will recognize all of the characters. The data taken does show, however, that the

recognition rates, even with varied fonts, are high enough that the unrecognized characters could be guessed from context. When the reading machine is built so that the reader can trim it up to the quality of the print being read it is expected that the recognition rates will go still higher.

In conclusion, it is the contour tracing with the incorporation of the large-threshold hysteresis smoothing which permitted a solution to this problem. The variability of the data after scanning is the cause of expensive data reduction in other techniques. Almost all the variability of the character is eliminated in the scanning process itself while maintaining the information describing the character. This is evidenced by the fact that the unknown characters can be described by 32 bits of information immediately after scanning and by the small amount of storage needed for the classification of characters. The simple operations used to generate the trace, the hysteresis smoothing technique used to reduce variability, and the small storage requirements suggest the real possibility of an inexpensive reading machine. It is estimated that a reading machine operating as described including all the optical scanning equipment could currently be built for a few thousand dollars, thus making a personal reading machine feasible.

BIBLIOGRAPHY

- (1) Coffee, J. L., "The Development and Evaluation of the Battelle Aural Reading Device," Proceedings of the International Congress on Technology and Blindness, the American Foundation for the Blind, New York, Vol. 1, 1963.
- (2) Troxel, D. E., "Tactile Communication," Ph.D. Thesis, M.I.T., August, 1962
- (3) Donaldson, R. W., "Multimodality Sensory Communication," Ph.D. Thesis, M.I.T., June, 1965.
- (4) Meffessel, M. F., "Experimental Studies of Human Factors in Perception and Learning of Spelled Speech," Proceedings of the International Congress on Technology and Blindness, the American Foundation for the Blind, New York, Vol. 1., 1963.
- (5) Shannon, C. E., "Prediction and Entropy of Printed English," The Bell System Technical Journal, Vol. 30, Jan. 1951, pp. 50-64.
- (6) Miller, G. A. and E. A. Friedman, "Reconstruction of Mutilated Texts," Information and Control, Vol. 1, No. 1, Sept. 1957, pp. 38-55.
- (7) Hannon, W. J. "The RCA Multi-Font Reading Machine," Optical Character Recognition, Sparton Books, Washington, D.C., 1962, pp. 3-14.

- (8) Chatten, J. B. and C. F. Teacher, "Character Recognition Techniques for Address Reading," Optical Character Recognition, Spartan Books, Washington, D.C., 1962, pp. 51-60.
- (9) Heasley, C. G. and C. F. Fisher, "Some Elements of Optical Scanning," Optical Character Recognition, Spartan Books, Washington, D.C., 1962, pp. 15-26.
- (10) Bomba, J. S., "Alpha-Numeric Character Recognition Using Local Operations," Proceedings of the Eastern Joint Computer Conference, Boston, Mass. Dec. 1-3, 1959, No. 16., pp. 218-224.
- (11) Rabinow, J., "Developments in Character Recognition Machines at Rabinow Engineering Co.," Optical Character Recognition, Spartan Books, Washington, D.C., 1962, pp. 27-50.
- (12) Greanias, E. C. et al., "The Recognition of Handwritten Numerals by Contour Analysis," IBM Journal of Research and Development, January, 1963.
- (13) Bradshaw, J. A., "Letter Recognition Using Captive Scan," I.E.E.E. Trans. on Electronic Computers, EC-12, February, 1963.

- (14) Bledsoe, W. W. and I. Browning, "Pattern Recognition and Reading by Machine," Proceedings of the Eastern Joint Computer Conference, Boston, Dec. 1-3, 1954, No. 16, pp. 225-232.
- (15) Rosenblatt, Frank, "The Perceptron," Psychological Review 65:6, Nov. 1958, pp. 386-404.
- (16) Eden, M., "Handwriting and Pattern Recognition," IRE Transactions on Information Theory, IT-8-2, February 1962, pp. 160-166.
- (17) Mermelstein, P., "Computer Recognition of Handwritten Words," D.S. Thesis, M.I.T., 1964.
- (18) Teager, H., "Multidimensional Visual Information Processing," Proceedings of the New York Academy of Science, New York, June 1965.
- (19) Earnest, L. D., "Machine Recognition of Cursive Writing," Information Processing 1962, Proceedings of the IFIP Congress, 1962, C. M. Popplewell, Ed., North Holland Publishing Co., Amsterdam, pp. 462-466.

BIOGRAPHY

Jon Kaufmann Clemens was born in May, 1938. He attended the Christopher Dock Mennonite High School in Lansdale, Pennsylvania and graduated in 1956. He then attended Goshen College, Goshen, Indiana in the 3-2 Engineering Program. After completing three years at Goshen College he transferred to the Massachusetts Institute of Technology. He received a Bachelor of Arts Degree in Physics from Goshen College in June, 1960, a Bachelor of Science and Master of Science Degree in Electrical Engineering from M.I.T. in June, 1963, and a Doctor of Philosophy Degree in Electrical Engineering from M.I.T. in September, 1965. During Graduate School at M.I.T. he was a Teaching Assistant for two years and a Research Assistant for two years. Summer employment positions were held at Digital Equipment Corp., Maynard, Mass.; Lincoln Laboratory, Lexington, Mass.; Bell Telephone Laboratory, Holmdel, N. J.; and Honeywell Corp., Waltham, Mass.

He is a member of the Tau Beta Pi and Eta Kappa Nu Societies and an associate member of the Sigma Xi Society.