Proof. This can be easily obtained by simple probabilistic considerations because p^*_{cj} is the probability that on leaving G_1 state V_j is reached.

Example 6. The values S_{ij} for a Markov chain M with a graph of Figure 3 can be computed in the following way:

(a) The values S'_{ij} for Markov chain M' with graph G' given in Figure 5 are computed. The labeling of the arcs of G' is identical with the labeling of the corresponding arcs of G except for arcs (V_6 , V_6) and (V_5 , V_5), which are labeled by 1.

(b) The values of S'_{26} and S'_{45} label the arcs (V_c , V_6) and (V_c , V_5) of the graph of Markov chain M^* with the graph given in Figure 4; the labeling of the remaining arcs is identical with the labeling of corresponding arcs in Figure 3.

(c) The values S^*_{ij} for M^* are computed.

(d) The values of S'_{ij} are multiplied by $p^{*\infty}_c$.

Note that if M fulfills the assumptions of Theorem 1, then M' and M^* must fulfill them as well, and, Theorem 1 is applicable to both M^* and M'.

Corollary 1. *As Theorem 4 is applicable to the one-entry subgraphs themselves the computation of the* S_{ij} *can be effected as follows:* (*1*) *The values* S'_{ij} *for the "innermost" subgraphs are computed.* (*2*) *These* S'_{ij} *are* (*in the sense of Theorem 4*) *used in the nearest "higher" subgraphs to compute* S^*_{ij} *for them until the* S_{ij} *in the highest* (*"outermost"*) *graph* G *are computed. Then using* (*5*) *of Theorem 4 the* S_{ij} *for inner subgraphs are successively computed.*

Remark 6. Treating "two-entry" subgraphs as two one-entry subgraphs, a variant of Theorem 4 for two-entry subgraphs (or n-entry subgraphs) can be proved.

Remark 7. Note that in graph G of an associated Markov chain the one-entry subgraph corresponds to a closed subroutine. This is the motivation of Remark 5 of Section 1. The details are not discussed here, but it can be shown that for the majority of subroutines the method indicated in Remark 5 practically does not increase the complexity of the computation of S_{ij}.

REFERENCES

1. Berge, C. *Théorie des Graphes et ses Applications.* Dunod, Paris, 1958.
2. Feller, W. *An Introduction to Probability Theory and Its Applications, Vol. 1.* Wiley, New York, 1950.
3. Král, J. To the problem of segmentation of a program. *Information Processing Machines.* Research Institute for Mathematical Machines, Prague, 1965.
4. ——. The formulation of the problem of program segmentation in the terms of pseudo-Boolean programming. Kybernetika 4, 1(1968), 6–11.
5. Martin, D., and Estrin, G. Models of computation and systems—evaluation of vertex probabilities in graph models of computations. *J. ACM 14*, 2 (Apr. 1967), 281–299.
6. Zurmühl R. *Matrizen,* 2 ed. Springer-Verlag, Berlin-Göttingen-Heidelberg, 1958.

Preliminary Investigation of Techniques for Automated Reading of Unformatted Text

George Nagy*
IBM Thomas J. Watson Research Center,
Yorktown Heights, New York

Methods for converting unstructured printed material into computer code are experimentally investigated. An operator-controlled mode, depending on human demarcation of the various regions of the page for guiding the scanner, is implemented by means of a joystick and a CRT display. This mode, for which some performance figures are obtained, is thought to be suitable for processing very complicated material, such as technical journals.

For simpler material, for instance the "claims" sections of patents, and in applications where the utmost accuracy is not necessary, an unsupervised mode is advocated. Here, the textual portions of the page are located during a rapid prescan by a rudimentary form of frequency analysis. These areas are then rescanned at a higher resolution suitable for character recognition. Error rates of the order of 0.1 percent are obtained in a simple problem involving photographs of telephone company meter boards.

Other matters related to the design of a general purpose page reader, such as the segmentation of printed text, the possibility of time-sharing the scanner, interactive man-machine operation, and the facsimile reproduction of illustrations, are discussed.

KEY WORDS AND PHRASES: pattern recognition, character recognition, text reading, information retrieval, unformatted text, operator-controlled reader, online reader, text-image discrimination, reading machine
CR CATEGORIES: 3.63, 3.79, 3.89, 6.29, 6.35

Introduction

The object of the project described here is to find convenient methods for subdividing a printed page into "read," "graphic," and "omit" fields in order to facilitate automatic page reading. Both operator-controlled and completely autonomous systems are considered. Even if it appears that in the foreseeable future man-machine

* Present address: Département d'Informatique, Université de Montréal, Montreal, Quebec, Canada.

systems offer the best hope for a "universal" page reader, automatic features would increase throughput and allow unsupervised operation on more constrained material.

Most character recognition machines now in use [1] require either a fixed format document, where the material to be read appears in well-defined fields, or manual editing with colored pencil or magnetic ink to outline the areas to be processed. More versatile readers would find application in massive file conversion and information retrieval projects, and in automatic translation, abstracting, and indexing.

In an *interactive* system it should be possible for the operator to direct the scanner only to areas of the page of interest by virtue of their meaning, and set in a typeface compatible with the recognition logic of the machine. "Graphics," including photographs, line drawings, graphs, charts, and esoteric symbols, could be scanned without any attempt at recognition, and stored in binary arrays for eventual redisplay by some type of facsimile device. Important headings, equations, footnotes, and other non-machine readable textual material would be typed in by the operator on an alphanumeric keyboard and stored in code. Alternatively, the operator would have the option of keying in a few lines of a new font in order to provide the machine with an identified training set to adjust its decision parameters. A function keyboard would allow labeling the various portions of the text to facilitate subsequent retrieval. Some preliminary experiments, designed to expose the problems likely to be encountered in implementing these ideas, are described in the next section.

Whenever the structure of the documents is sufficiently simple and uniform to allow completely *automatic* processing, there is much to be gained from this mode of operation. Several experiments in automatic page decomposition were carried out on material ranging in complexity from photographs of telephone company meter boards to technical journal pages. The method used is based on a rudimentary form of spectral analysis with a word size scan window.

The IBM experimental character recognition system on which these experiments were performed already incorporates a number of fairly sophisticated features for registration, normalization, noise suppression, threshold adjustment, and character separation. When presented with a noncharacter field, it attempts to convert what it finds into familiar symbols by means of protracted and agonizing convulsions. What is needed, then, is a rapid "prescan" of the entire page, or a sizable fraction thereof, to indicate to the control unit which areas are to be read, and which ones skipped or copied onto magnetic tape without further processing. The logic used to derive this information from the prescan is discussed in the third section of this report.

Operator Controlled Scan

To minimize design and construction time, and to obtain preliminary results as rapidly as possible, the existing opaque page reading facility at the IBM Thomas J.

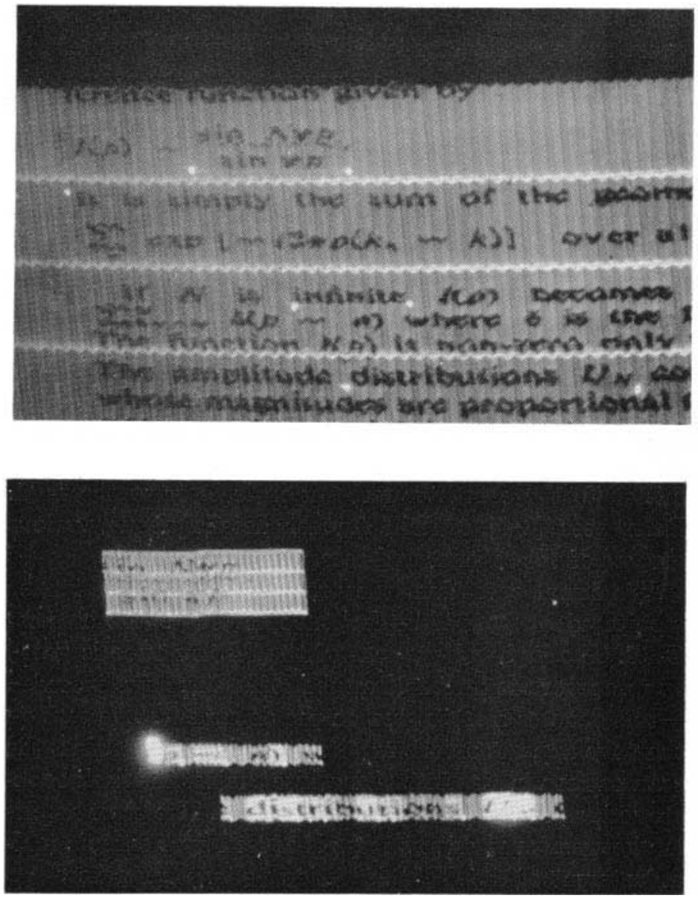


FIG. 1. Operator-controlled raster scan
TOP: Illumination pattern to allow operator to read and mark text
BOTTOM: Process mode. The formula is stored in facsimile form, while the two words are segmented and the letters individually recognized. The bright areas show letters which were rescanned with altered parameters, as a result of rejection by the decision unit, to improve their binary representation.

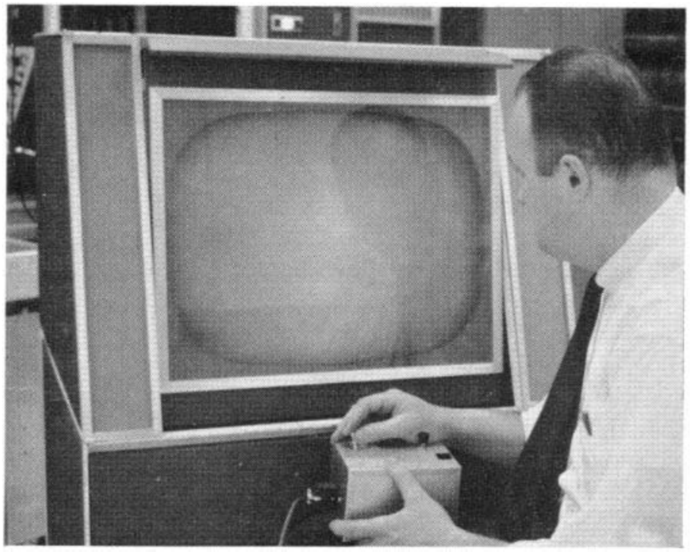


FIG. 2. The joystick control
The operator faces the screen and directs the CRT spot by means of a joystick connected to the resistors regulating the repetition frequency in two pulse generators feeding the X and Y registers.

Watson Research Center was converted for use in the operator controlled mode. Although this arrangement is, as expected, inconvenient in many respects, sufficient data has already been collected to justify investment in more elaborate hardware and software.

The scanner in the existing system comprises a cathode ray tube, optics to focus the light from the flying spot onto the document, and a light collection system with eight photomultiplier tubes. The deflection circuits are controlled by a special purpose digital-to-analog interface which receives macro-commands, such as "search," "read," or "increase vertical raster resolution," from an IBM 1401 computer. After compensation for phosphor noise the signal from the photomultiplier tubes is thresholded, buffered in a 36-bit register, and finally transmitted to the character recognition logic or to the 1401. References [2–4] contain further details about various aspects of the scanner and recognition system.

In normal operation, the progress of the spot can be monitored on a 21″ display tube slaved to the scanner CRT. The high-voltage gun of the display tube is gated *on* whenever the threshold on the PMT current is exceeded —this denotes a *white* area on the document. The text on the page appears on the monitor tube in black against the scan pattern, which in turn is white against the dark background, as shown in the top of Figure 1. The zigzag overlap between successive sweeps is caused by a chronic maladjustment of the CRT deflection voltage generator.

Display and Joystick Control. As has been shown above, in order to take advantage of the capabilities of a human operator, he must be provided, in addition to a display of the document being processed, with some means of directing the scanner to specific areas of the document. With roughly 1000 × 1500 bits necessary for legibility on an $8\frac{1}{2}''$ × 11″ page with normal print density, a display regenerated from the core memory of the 1401 was clearly out of the question. Neither was a suitable storage tube readily available. Instead, the relatively long persistence of the monitor tube is taken advantage of to provide a stable image for the operator. When an $8\frac{1}{2}''$ × 3″ portion of the page is scanned in 1.8 seconds at a resolution of 160 lines per inch, the resulting image remains legible for about 6 seconds. This time can then be used by the operator to perform the "editing" function.

The operator's principal mechanism of control is the "joystick" shown in Figure 2. The joystick is used in a "rate control" mode. The direction in which the stick is tilted determines the direction in which the spot moves, and the velocity of the spot motion depends on how *far* the stick is tilted. While a lightpen, or even a joystick with a "proportional" mode of control (where the position of the spot is determined directly by the position of the stick), would be clearly preferable, in the existing hardware the necessary feedback loops would have required a disproportionate amount of modification.

In addition to the joystick, the operator is provided with a pushbutton which causes the position of the spot to be registered by the 1401 as soon as it has been moved to the right place, and with several function switches which allow him to specify how the information in a particular region is to be processed.

Here is a typical sequence of operations. The spot appears on the face of the monitor tube. By means of the joystick, the operator moves the spot to a region of interest. A raster scan pattern "illuminates" the portion of the page centered about the last position of the spot, allowing the operator to underline certain words ("infinite" and "distributions" on Figure 1) and flag them for character recognition, and to underline others ("sin$N\pi p$/sinπp" on Figure 1) and flag them for "facsimile" storage by means of the function switches. At the end of six seconds control of the spot is taken away from the joystick and returned to the computer and the raster scan mode is used again to illuminate a portion of the page centered about the last position of the spot. JOYSTICK and ILLUMINATE modes alternate until the operator depresses another function switch: the scanner now goes into PROCESS mode (bottom of Figure 1). The areas flagged for character recognition are rescanned at higher resolution, and the recognized identities of the characters are stored on magnetic tape. The areas flagged for facsimile are also scanned at a preset resolution, and the video information is stored as well as printed out (Figure 3).

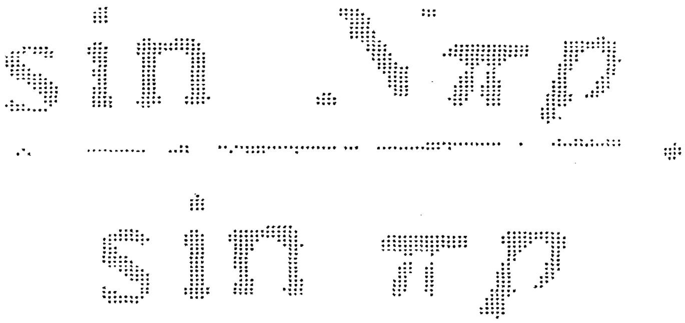


FIG. 3. Facsimile representation

Binary video printout of a formula, in the form it would be stored for eventual redisplay. Improvements are needed in terms of higher resolution and the introduction of grey levels.

The time required to process a single page is, of course, dependent on: the amount of material flagged; the difficulty of the criteria used by the operator for flagging a word, sentence, or picture; the quality of the display; and the ease of entering into the system the operator's decisions. To estimate the relative importance of some of these variables, and to provide bounds on the performance level attainable by a practical system, a few simple experiments were conducted.

Experiments

The object of the first experiment was to set up a standard against which display systems and control mechanisms could be judged. Columns 1, 2, 3, and 4 of Patent No. 2,816,237, "System for Coupling Signals Into and Out of Flip-Flops" were used. The operator simply underlined with a pencil every word on the two pages in question; this took 4.1 and 4.3 minutes respectively. Thus about *360 words per minute* represents an upper bound on the speed attainable in a mode where the operator has to mark every word to be scanned.

In the second experiment the same operation was performed using the monitor display tube and the joystick control. Slight differences in timing between two operators and two different illumination cycles altering the time between successive displays are shown in the following table.

	Illumination		*Total time*	
Material	*Cycle time*	*Duration time*	*Operator 1*	*Operator 2*
	(in minutes)		(in minutes)	
Columns 1, 2	.10	.03	23.1	24.3
Columns 3, 4	.20	.03	18.5	21.2

After making allowance for the duration of the ILLUMINATION mode during which the operators were unable to control the spot, an average speed of *81 words per minute* is obtained. The difference between this figure and the one above is directly attributable to the shortcomings of the joystick and its peculiar derivative mode of control.

The percentage of correctly identified letters in the words underlined varied from 95-98 percent from day to day depending on the care taken to adjust the video circuits of the scanner. This includes only lowercase characters, since there was no recognition software available for capitals, punctuation, and ligatures (fi, ff, fl, ffl, and ffi).

Very little effort was expended in developing a recognition logic suited for the print style used in patent documents. The measurements (features) normally used with typewritten material were taken over without change, and the decision parameters of a simple linear categorizer were determined by means of a small sample whose identities were keypunched.

A third experiment was designed to test the operator's ability to cope with formulae, esoteric symbols, and italics in the text, and other impediments to automatic recognition. On several typical pages of the *IBM Journal of Research and Development* [Volume 7, No. 4, October 1963] the operators were asked to underline all potentially troublesome material. The material on page 219 is shown on Figure 4; it is seen that due to the poor quality of the display many (almost a quarter) of the "obstacles" were missed. The relative speed and performance of the two operators on this and other pages appear in the table below.

	Number of "Omit" fields		*Total time*	
Material	*Operator 1*	*Operator 2*	*Operator 1*	*Operator 2*
page 279	23	6[a]	7.5	4.3
page 281	28	20	8.5	6.2
page 296	15	13	4.9	4.9
page 310	27	25	7.3	9.3

[a] Inexperienced operator

Of course, for purposes of automatic abstracting and indexing, missing the occasional strange symbol embedded in the text would hardly cause serious difficulties.

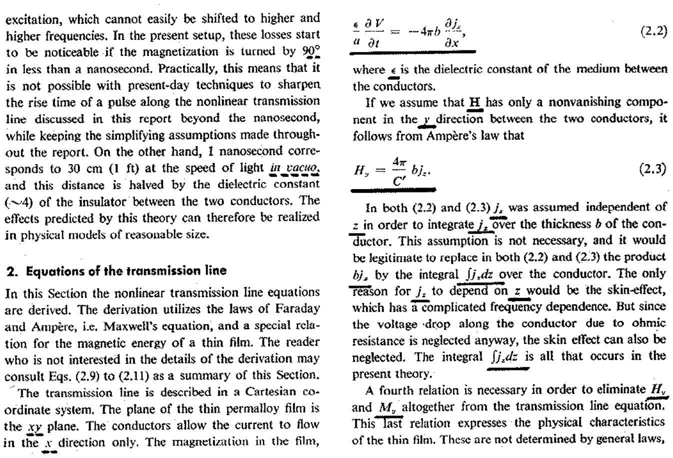
Automatic Page Decomposition

The method to be discussed stems from the simple observation that character fields are readily distinguishable from almost everything else by the average density of the lines and of the blank spaces above and below the lines.

A few hours in a darkroom will suffice to demonstrate this hypothesis; accordingly, a simple photographic experiment will be described first. Of course, the photographic process is not really practical for a high speed page reader, but direct optical implementation may be devised. The next step was digital computer simulation. Simulation on the IBM 1401 proved to be very slow, but showed enough promise to warrant hardware implementation. The final system, with a hardware "word locator," was tried on photographs of telephone company meter boards obtained from the German PTT (Post Telephone and Telegraph) authority.

Optical Masking. The aim here is to prepare a transparency mask whose clear areas correspond to the text. The process is as follows:

First, an intermediate mask, consisting of a highly defocused negative of the page at 1:1 magnification, is obtained. Such an intermediate mask is shown in the center of Figure 5, for the page on the left-hand side. Were this mask used directly to photograph the page, the result would contain many damaged characters at the beginning and end of words. In addition, isolated letters are occasionally lost, and not all the graphic information is suppressed. Some improvement may be obtained by preparing a final mask by exposing horizontal and vertical translations of the intermediate mask. The intermediate



excitation, which cannot easily be shifted to higher and higher frequencies. In the present setup, these losses start to be noticeable if the magnetization is turned by 90° in less than a nanosecond. Practically, this means that it is not possible with present-day techniques to sharpen the rise time of a pulse along the nonlinear transmission line discussed in this report beyond the nanosecond, while keeping the simplifying assumptions made throughout the report. On the other hand, 1 nanosecond corresponds to 30 cm (1 ft) at the speed of light *in vacuo*, and this distance is halved by the dielectric constant (~4) of the insulator between the two conductors. The effects predicted by this theory can therefore be realized in physical models of reasonable size.

2. Equations of the transmission line

In this Section the nonlinear transmission line equations are derived. The derivation utilizes the laws of Faraday and Ampère, i.e. Maxwell's equation, and a special relation for the magnetic energy of a thin film. The reader who is not interested in the details of the derivation may consult Eqs. (2.9) to (2.11) as a summary of this Section.

The transmission line is described in a Cartesian coordinate system. The plane of the thin permalloy film is the xy plane. The conductors allow the current to flow in the x direction only. The magnetization in the film,

$$\frac{\epsilon}{a}\frac{\partial V}{\partial t} = -4\pi b \frac{\partial j_x}{\partial x}, \qquad (2.2)$$

where ϵ is the dielectric constant of the medium between the conductors.

If we assume that $\mathbf{H}$ has only a nonvanishing component in the y direction between the two conductors, it follows from Ampère's law that

$$H_y = \frac{4\pi}{C'} b j_x. \qquad (2.3)$$

In both (2.2) and (2.3) j_x was assumed independent of z in order to integrate j_x over the thickness b of the conductor. This assumption is not necessary, and it would be legitimate to replace in both (2.2) and (2.3) the product bj_x by the integral $\int j_x dz$ over the conductor. The only reason for j_x to depend on z would be the skin-effect, which has a complicated frequency dependence. But since the voltage drop along the conductor due to ohmic resistance is neglected anyway, the skin effect can also be neglected. The integral $\int j_x dz$ is all that occurs in the present theory.

A fourth relation is necessary in order to eliminate H_y and M_y altogether from the transmission line equation. This last relation expresses the physical characteristics of the thin film. These are not determined by general laws,

FIG. 4. Test of operator performance

Portions of the material which were labeled by the operator to be omitted in automatic scanning are underscored with solid lines. Dotted lines indicate symbols that he missed.

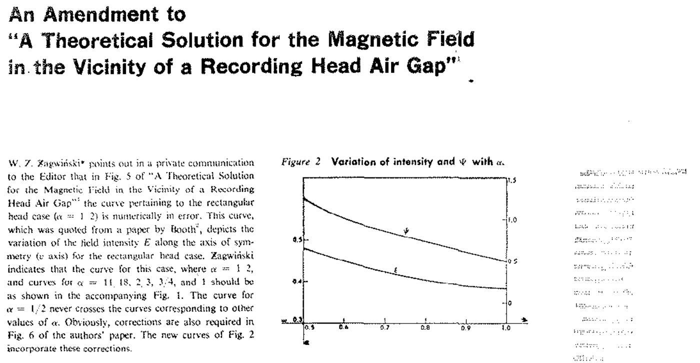
mask is slowly moved about during exposure by an amount corresponding to about five character widths in the horizontal direction, and half a character height in the vertical direction. The final copy, which does emphasize text at the expense of the illustrations, is shown on the right of Figure 5.

Computer Simulation. The present scanning facility offers no provision for reregistering a document once it has been removed from the scanner bed. Thus, short of scanning the whole page at high resolution and doing the character recognition as well as the page decomposition on a large computer, all of the processing had to be carried out on the IBM 1401 which controls the scanner.

Because of the limited storage capability of the 1401, the calculation of the local average density of the different portions of the page is carried out by the WINDOW routine as the scanner progresses down the page in the PRESCAN mode. The TEXTMARK subroutine then checks certain other simple conditions and writes out the coordinates of the text portions of the page in the form of a "T-tape." These areas are then rescanned in the CHARACTER RECOGNITION mode.

During PRESCAN, the scanner converts a horizontal strip of the printed page into a 32 × 1000 matrix of black and white points. The vertical resolution is set equal to the horizontal resolution, so that an $8'' \times 11\frac{1}{2}''$ page yields about 1,300,000 bits. Successive strips are scanned and processed by the WINDOW subroutine until the whole page is covered.

The WINDOW subroutine calculates the number of black bits in a set of overlapping rectangular windows. The window is 80 bits wide and 20 bits high, approximately corresponding in size to a five-letter word. The overlap is about 75 percent in both directions, causing four times denser coverage vertically than horizontally.


An Amendment to "A Theoretical Solution for the Magnetic Field in the Vicinity of a Recording Head Air Gap"[1]

W. Z. Zagwiński* points out in a private communication to the Editor that in Fig. 5 of "A Theoretical Solution for the Magnetic Field in the Vicinity of a Recording Head Air Gap"[1] the curve pertaining to the rectangular head case ($\alpha = 1/2$) is numerically in error. This curve, which was quoted from a paper by Booth[2], depicts the variation of the field intensity E along the axis of symmetry (v axis) for the rectangular head case. Zagwiński indicates that the curve for this case, where $\alpha = 1/2$, and curves for $\alpha = 1/1$, 18, 2/3, 3/4, and 1 should be as shown in the accompanying Fig. 1. The curve for $\alpha = 1/2$ never crosses the curves corresponding to other values of α. Obviously, corrections are also required in Fig. 6 of the authors' paper. The new curves of Fig. 2 incorporate these corrections.

Figure 2 **Variation of intensity and ψ with α.**

FIG. 6. Automatic page decomposition

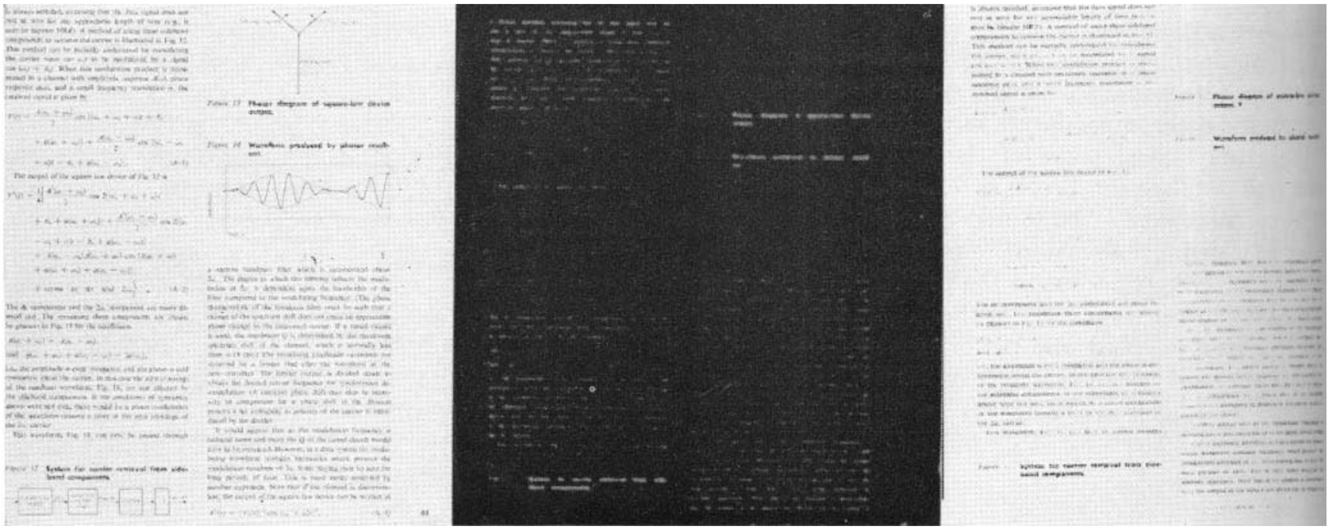
LEFT: Intermixed text and figures

RIGHT: T-tape generated by IBM 1401, showing text areas to be scanned at high resolution. Some of the equations are also suppressed, as is the heading. The proportions of the page are distorted due to the low horizontal resolution used in the prescan mode.

The TEXTMARK subroutine checks whether a particular window fulfills the density conditions specified for the font under consideration and whether it has sufficient light (white) areas above and below it to qualify as text. If only the latter condition is met, a "conditional textmark" is set and windows to the right and left are then examined to see if they form part of the text. A conditional textmark is transformed into an unconditional textmark if a textmark occurs on either side of it.

The OR-ing process is designed to recover isolated letters and symbols. Figure 6 shows the textmark fields on the T-tape generated by this program. About seven lines on the printout correspond to one line of text.

For the purpose of this experiment the CHARACTER RECOGNITION routine was modified to scan and process


A page from the *IBM Journal*

Intermediate mask generated from defocused image of the page on the left

Final copy of document photographed through translated mask

FIG. 5. Optical masking

only the areas of the page corresponding to the textmarks. An addition to the SEGMENTATION routine was helpful in dealing with spurious textmark fields caused by horizontal picture elements: whenever forced separation had to be employed more than twice in a row, the whole textmark field was rejected.

The running speed of the WINDOW and TEXTMARK programs was about 90 minutes per journal page, consequently a simple hardware implementation was sought.

Hardware Implementation. The central item in the hardware implementation is a long (780 bit positions) shift register which is also used to provide registration invariance for character recognition. Ones and zeros from the scanner, corresponding to black and white areas of the image, are piped to the shift register, where the bits become available for various logical operations. Because the scanner progresses across the page in 32-bit vertical scans, with a fixed number of blank dummy bits between each scan, the image in the shift register may be visualized as spiraling around a cylinder (Figure 7).

In order to qualify as text, a given portion of the image must contain only white bits near the top and bottom, and a certain density or proportion of black bits near the center of the raster. This condition may be detected by means of the logic shown in Figure 7.

The white condition requires only that a forty input AND-gate (made up of a number of cascaded three-way AND's) be satisfied. The grey condition is a little more difficult, since which bit positions are "on" depends on the text being scanned. Here an analog adder is used, with two threshold circuits which are simultaneously satisfied only if between 30–40 percent of the bits in the central area are "on."

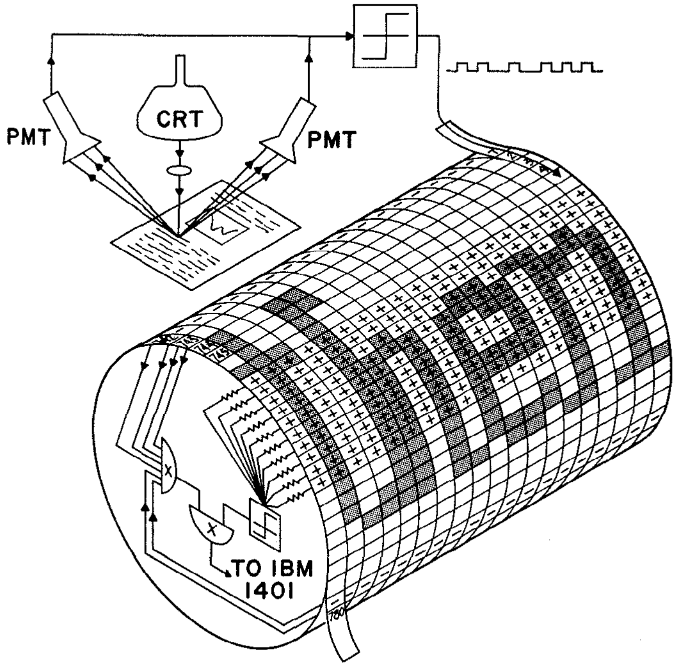


FIG. 7. Shift register and window logic

The 780 bit shift register is represented as a spiral wrapped around a cylinder. Shift register positions where the logic is connected to the positive output (black) are marked "+", while positions tapped on the negative output (white) are marked "−". The word "then", scanned at reduced horizontal resolution, is shown shifting through the register. Whenever the final AND-gate is satisfied, a conditional text-mark is set.

The vertical resolution is still set so that a capital letter is about 21 bits high, but the horizontal resolution is decreased sufficiently to allow a five-letter word into the shift register at one time. The final AND-circuit is interrogated every five scans; if it is "on," a textmark is set. The OR logic, to recover isolated characters, is still performed by the 1401 computer, although direct electronic implementation would be simple. The remainder of the recognition process is executed as before, with the text marks guiding the scanner to the appropriate areas of the page.

The Meter Board Problem. In Germany the number of message units expended by customers of the PTT appears on mechanical counter arrays, as shown on the left in Figure 8. The PTT would like to automate the transcription operation, which is presently performed by keypunchers. The system discussed in the previous section was used to provide a demonstration of the feasibility of using automatic character recognition equipment to per-

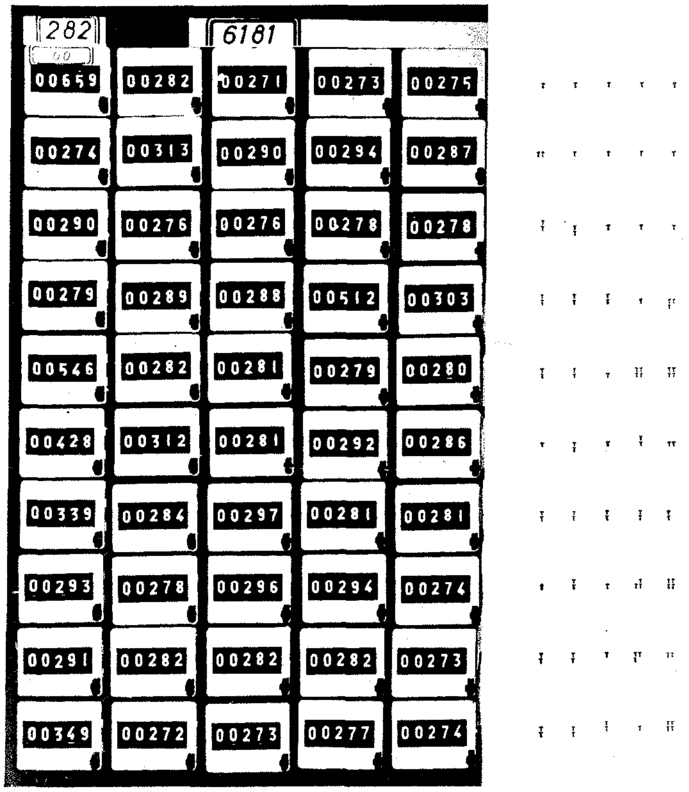


FIG. 8. Reading meters for the telephone company

LEFT: Meter board array with 50 counters.

RIGHT: T-tape generated with special purpose logic shown on Figure 9. The counters are located sufficiently accurately to allow the local registration scheme in the RECOGNITION routine to take over.

form this function. A fixed format procedure might have difficulties with this problem because the relative position of the meters may vary by as much as a full meter height due to the cumulative displacement of the individual units.

To locate the meters on the photographs, and to avoid scanning the whole photograph at high resolution, a special form of the word-window, shown in Figure 9, was used. A T-tape written with this logic is shown on the right in Figure 8. The numerals in the windows located by the T's were scanned at a resolution corresponding of 125 lines per inch, and identified by means of a simple mask-matching program.

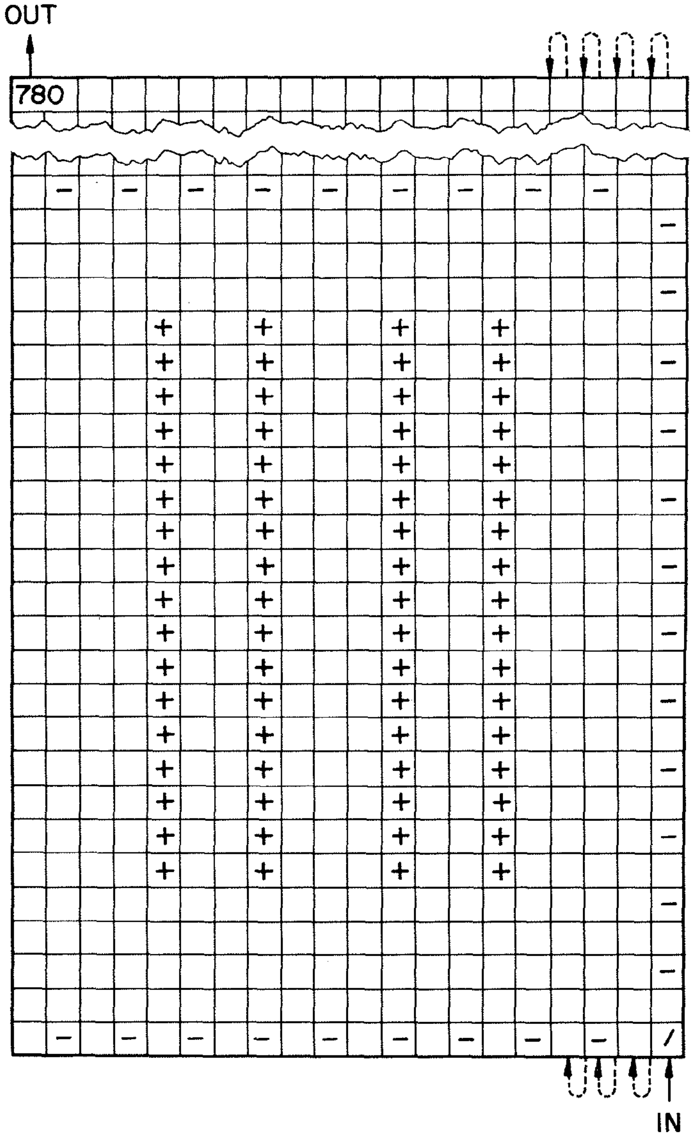


Fig. 9. Special WINDOW Logic for Counter Array
This logic evolved through trial and error from the general purpose array shown on Figure 7. The logic was modified whenever a window was missed or one erroneously signaled. The black bands between the numerals were of considerable help in accurate horizontal registration. Several styles of meters had to be accommodated.

On 5480 characters 99 percent recognition accuracy was achieved, with 0.90 percent rejects; thus only five characters were actually misrecognized. Four windows out of 1100 were missed by the counter locator.

The running time of the recognition program on an IBM 7094 computer was 32 characters per second. It is estimated that with additional shortcuts this speed could be almost doubled. On a system 360 model 40 or 50 computer, further gains in performance could be obtained by using to advantage the micrologic provided in the read-only memory.

Conclusions

The experience gained in the course of the studies outlined above suggests the following conclusions and recommendations.

Very simple, automatic methods are sufficient to decompose a printed page into *text* set in one, or a few, predetermined type styles, and *other material.* While there appears to be no convenient standard for measuring the accuracy of the decomposition, it would seem that more complicated and time-consuming methods of spectral analysis need not be invoked for this purpose.

With documents of the order of complexity of technical journals, automated scanning at this level, coupled with existing character recognition devices, would not produce a computer storable reproduction suitable for all of the intended purposes of the original document. A limited range of functions, such as automatic indexing, and perhaps extraction, could be performed on the coded version. Simpler page structures, such as the "claims" section of patent documents, could be processed for natural language search procedures and complete redisplay.

Operator guided scan patterns allow complete reproduction of the document. Considerable savings in computer storage space (of the order of a factor of 100 on a patent claim) result from coding the main body of the text, or selected portions thereof, by automatic character recognition techniques. In this mode of operation the operator himself can perform certain editing functions, including indexing and labeling. With a convenient display system and means of directing the scanner, processing speeds of 15 words per second can be expected on high-density information bearing text.

A relatively low resolution display, of the order of 150 lines per inch, is sufficient for most documents. Nevertheless, scanning the entire page by means of a flying spot scanner, and either recirculating the display from a buffer, or holding it on a storage tube, do not appear to be the most practical ways to proceed. The document itself, or a copy of it, would constitute the most economical form of storage. The document could be projected from the scanner bed through a dual optical system onto a RAND tablet or other X-Y device. Alternatively, a copy of the document could be superimposed on the tablet. The registration problem is easily surmountable, since registration within $\frac{1}{8}''$ is ample.

Because the most expensive portions of character recognition systems are traditionally the document transport and the scanner itself, it would make sense to time-share these components whenever offline scanning is practicable. Even a very small computer could keep house for several RAND tablets and operators. The areas of interest on each document are delineated with the stylus and labeled with function switches, and the coordinates and labels are stored on disk or tape file. In a subsequent operation, the documents are registered in the scanner, and the operations specified on the file performed. At 10 seconds per page per operator, and current character recognition and scanning speeds (1000 characters per second or 1,000,000 bits per second of video), one flying spot scanner could keep pace with about 12 operators on standard sized journal pages consisting of 50 percent text and 50 percent illustrations.

New Equipment. In order to test the ideas discussed in a more realistic environment, as well as to conduct research in other aspects of pattern recognition, a new scanning facility is in the process of being installed at the Thomas J. Watson Research Center. The pertinent portions of the system comprise an IBM 1800 computer with extensive interrupt and analog-digital capabilities in overall control, a cathode-ray tube scanner, RAND tablet, hardware for character recognition logic, a 21″ display tube, and peripheral equipment. This configuration is presently at the debugging stage, with the basic support software well underway.

Aside from the improvement over the joystick represented by the RAND tablet, programming is simplified by direct control of the CRT spot by the 1800, without a special purpose interface.

Future Activities. One area of paramount importance, which has been neglected in the present study, is the recognition and segmentation of printed text. The 98 percent recognition on patent claims, which is the best obtainable with the existing system, is surely unnecessarily low for most applications.

The segmentation routines developed for touching characters in typed text are not applicable to variable pitch print, where the width of the letters may vary in a 5:1 ratio (w:i). Although there is reason to believe that if the segmentation problem were solved recognition performance similar to that obtainable on single font typewritten material could be achieved, this remains to be shown. It is possible that on some printed material the segmentation can be carried out only by means of a floating recognition scheme, which processes longer than letter length segments of video. Additional difficulties are anticipated with the large alphabet (of the order of 110 symbols) necessary for print reading; the typecase includes many difficult punctuation symbols.

The basic recognition accuracy can be greatly enhanced by including in the system some of the more sophisticated methods developed in the laboratory in the last few years. These include the use of context in both backward- and forward-looking modes [5], taking advantage of the correlations between "measurements" by means of nonlinear decision surfaces [6], the use of tracking algorithms in a "self corrective" mode [7], and the even more massive recognition schemes based on higher order autocorrelations [8]. In spite of the formidable programming task represented by adding these functions to the already overburdened scan, display, interact, and basic recognition programs, these refinements may be necessary for obtaining acceptable error rates on the large amounts of heterogeneous data needed for realistic evaluation.

Another area which would bear investigation is that of higher level interaction between the control computer and the operator to permit even greater variability in the input documents and flexibility in the output. The scanner could proceed autonomously until it encounters "unreadable" material, which it then displays on a screen for the operator's attention (a far smaller display than that necessary for an entire page would be sufficient here). The operator then simply indicates to the scanner what action to take: to scan the material, as before, in a facsimile mode; to resort to curve following; to summon its arsenal for italics, boldface, or superscripts; or to let the operator key in the offending word or letter. Aside from the intrinsic economies which may be realized on some classes of character recognition applications, the experience gained here may be useful in other man-machine interaction situations.

Acknowledgments. The computer program and experimentation were executed by Charles Marr, with the assistance of Morris Grove, Martha Miller, and Arthur Sebastiano. Peter Will and Ed Carey designed the joystick circuits and logic, and Don Fraleigh assembled the hardware for the "window." John Harper performed the optical work. The considerable experience of Glen Shelton, Jr., was freely drawn upon.

RECEIVED JULY, 1967; REVISED NOVEMBER, 1967

REFERENCES

1. FEIDELMAN, L. A. A survey of the character recognition field. *Datamation* (Feb. 1966), 45–52.
2. POTTER, R. J. An optical character scanner. *J. Soc. Photogr. Instrum. Eng. 2*, 3 (Feb. 1964), 75–78.
3. LIU, C. N. A programmed algorithm for designing multifont character recognition logics. *IEEE Trans. EC-13*, 5 (Oct. 1964), 586–593.
4. LIU, C. N., AND SHELTON, G. L., JR. An experimental investigation of a mixed-font print recognition system. *IEEE Trans. EC-15*, 6 (Dec. 1966), 916–925.
5. RAVIV, J. Decision making in Markov chains applied to the problem of pattern recognition. IBM Rep. RC 1672. Yorktown Heights, N.Y., Aug. 1966. Also *IEEE Trans. IT-13*, 4 (Oct. 1967), 536–551.
6. CHOW, C. K., AND LIU, C. N. An approach to structure adaptation in pattern recognition. Proc. 1966 Nat. Electron. Conf., Vol. 22, pp. 573–578.
7. NAGY, G., AND SHELTON, G. L., JR. Self-corrective character recognition system. *IEEE Trans. IT-12*, 2 (Apr. 1966), 215–222.
8. MCLAUGHLIN, J. A., AND RAVIV, J. *N*th order autocorrelations in pattern recognition. Proc. 1967 Int. Symp. Inform. Theory. In press. Also IBM Rep. RC 1790, Yorktown Heights, N.Y., Mar. 1967.