PROSPECTS AND PROBLEMS IN DESIGNING IMAGE ORIENTED INFORMATION SYSTEMS†

S. R. Ray

Introduction

There are slowly maturing and growing about us today a number of techniques which are likely to have a very significant effect upon the implementation of information systems in the near future. One of these techniques is pictorial data handling and interpretation, which is a subclass of the general area called pattern recognition. Pictorial data processing first became volumetrically significant in the case of photographic output of synchrotron bubble chambers which now deliver several million photographs per year.[2] More recently, a surge of interest has developed in automatic interpretation of biological, medical, and weather satellite pictorial data. The automatic scanning of microscopic slides for the purpose of identifying certain morphological characters is an example of a rather complex task in the area of biological/medical laboratory automation.

Some new viewpoints have begun to emerge from the experience of grappling with large volume pictorial data handling problems.

First, consideration of these problems has led to innovations in computer organization. Let me expand briefly upon this point. The usual organization of a digital computer is strongly oriented toward arithmetic calculation; hence, a reasonable unit of information is some 6 to 12 decimal digits and this unit (some 20 to 40 bits), a word, is all that can be processed in a single operation. On the other hand, a unit of pictorial information may consist of a million or so bits

---

†Supported in part by Contract AT(11-1)-1018 with the U. S. Atomic Energy Commission and the Advanced Research Project Agency. Many valuable suggestions have been contributed by B. H McCormick (see footnote 1) and R. Amendola. The original manuscript was prepared by Anita Worthington and John Otten.

S. R Ray is Associate Professor of Electrical Engineering, Department of Computer Science, University of Illinois, Urbana.

which makes profitable a very large information unit. Furthermore, the arithmetic word contains only one-dimensional positional information (the position weights of digits) whereas the pictorial word contains two-dimensional information (relative geographic position). Thus, pictorial data processing has led to the design of two-dimensional word processors which have data rates much higher than previous computers because of the large words processed in a single operation (e.g., words of 1024 bits).

Second, attacks on the problem of processing pictorial data originating from remote laboratories have focused increased attention on the video transmission and remote terminal facilities required to transmit high-quality, high-data-rate pictures. A concomitant requirement is that of permanent recording of pictorial data and its temporary display, storage, and copying at remote locations.

How can the foregoing developments be profitably applied to the problem of mechanized information retrieval? The remainder of this paper addresses the application of these concepts and techniques to an experimental ISRD (Information Storage, Retrieval, and Dissemination) system wherein economic balance among the storage, indexing and dissemination functions is attempted.

General Description of the System

Before proceeding to the important question of the basic premises of the proposed ISRD system, it is desirable to enumerate the major hardware components and sketch the operation of the system.

Beginning with the hardware system* in Figure 1, the centrally located equipment is clustered conceptually about the central processor. This device can be regarded as a digital computer with special capability for handling non-numerical data and, especially, for switching pictorial images.

A second centrally located device is the parallel processor, a unique device for analyzing and dissecting two-dimensional picture data rapidly. In the present context, the parallel processor is used for character recognition and page format analysis as described in detail subsequently.

The third hardware object is a microimage store which is simply a high-volume, fast-access store of data in image form (as opposed

---

*The hardware system described here is essentially the ILLIAC III. All major equipment was either in construction, on order from commercial sources, or in the design stage at the Department of Computer Science, University of Illinois, Urbana, in early 1967.

to bit-coded or Hollerith data form). For present purposes, the image store contains the entire document file.

The fourth hardware object is the bit store which is a high capacity, semirandom access magnetic disc or card store. This store contains index information relating to the document data in the microimage store.

The fifth major unit of equipment is a group of consoles with visual display, typewriter, light pen and assorted controls for manipulating data and performing command/control functions for the remaining equipment.

The final group of equipment in the central area consists of program-controlled flying-spot scanners capable of digitizing various sizes of optical film. In particular, one scanner accepts microfiche cards.

Remotely located stations communicate to the central equipment via a video network. A remote station itself contains a display facility as well as a recording mechanism capable of making hard-copy records of the displayed information.

Let us turn next to an outline of the operational system. In its infancy, the system will be operated in a man-controlled, machine-assisted mode which evolves toward greater automaticity as programming matures.

The storage operation begins with printed hard-copy material which is reduced first to primary microfiche, either locally or at some service facility. A small fraction of the primary microfiche images are reformatted onto secondary microfiche, after which all images are reduced to high-density microimage cards for the image store.

After insertion of the daily acquisitions in image store, rudimentary indexing operations are performed. In the immature system, indexing will be strongly controlled by an operator through the local control/display console. The operator will flip through pages in the image store, point with his light pen to (say) title and author data, which are then character-recognized (in the parallel processor) and, eventually, inserted in the index.

As the system matures, an increasing fraction of the indexing-object location task will be relegated to programmed procedures. More thorough indexing may also be performed on an increasingly automatic schedule, relying upon the full text image which remains available.

Finally, retrieval and dissemination proceed from the remote stations where a general user may query the index, obtain summary roadmap diagrams, browse through full text pages, etc. Permanent individual copy may be made locally in the form of microfiche employing the polaroid process.

Premises and Objectives

The fundamental premise of the ISRD system under discussion is that rapid accessibility and dissemination of information are essential features, as well as increasingly vital factors in total information service. The concept of direct user interaction, expounded in the INTREX report[3] and by Licklider,[4] for example, explicitly require on-line accessibility and rapid dissemination as a basic ingredient. In the present instance, we are not so greatly concerned with direct intercommunication between active users as with rapid accessibility and dissemination of data from a central location to users.

"Rapid dissemination and accessibility," as construed here, mean that, ideally, all documents of interest to any user shall be available in electronically accessible and transmittable form. More concisely, the objective is to design a system which, under realistic economic constraints and using present technology, approaches satisfaction of the "all document" accessibility-dissemination criterion.

The first crucial economic problem encountered is the cost of introduction of, perhaps, two to five million pages of original data per year. We propose to enter data into the system in image format only and the arguments for this approach follow:

> The cost of data introduction, per se, in image format is certainly the most economical technique presently feasible, besides being the fastest, especially in view of the fact that the data of interest is usually already in image format.
>
> Furthermore, the cost of a large-volume image store (e.g., 50 to 100 million pages) is less than that of a volumetrically comparable bit store.

Moreover, even a cursory examination of the operational requirements of an information system reveals that there is only one region in the system where information must be in Hollerith (i.e., bit) format. That region is the index and the reason for Hollerith format there is the practical efficiency of searching highly compact bit data as compared to loosely packed, relatively ambiguous image data.

The objective is to derive Hollerith data for the index from the stored image data only when it is needed. Thus, all documents entering the system are indexed crudely, for example, using title word descriptors and author names. As the demonstrable worth of a document grows, it is subjected to deeper indexing requiring conversion of more extensive portions to Hollerith format. In any case, only a relatively tiny fraction of all introduced text need ever be converted to Hollerith data. We estimate, for example, that if every document entering the system is indexed by only 100 characters and 2 percent of the documents are indexed on the basis of 10,000 characters, then only 12 days per year or 3 percent of the recognition system's operation time will be consumed in character recognition (at 1000 characters per second) for a five million pages per year entry rate.

Another major objective of the system is an evolutionary development toward highly automatic indexing capability with concomitant improvement in operational economy. The special problem of the image system is automatic geographical location of text objects for use in the index. An example of automatic location, that of finding figure captions, is given in a subsequent section.

At the present time, relatively little definitive work has been directed toward automatic location of objects in text, per se, although a substantial effort has been applied to related areas[5,6] for the problem is common to all pictorial data analysis. By extrapolation from picture processing experience, it is certain that repetitive, frequently recurring cases of text objects can be automatically located economically, i.e., considering both programming and operating costs. It is equally certain that error-free automatic handling is unachievable because of obscure and bewildering cases which must be treated in the manual back-up mode, i.e., by a console operator. Nevertheless, even very minimal constraints on printing formats would go far toward improving location accuracy and efficiency. Automatic location is performed largely in the parallel processor which, as suggested above, is quite incompletely occupied by character recognition. The same unit is also used for format analysis which is explained further on. Finally, it should be noted that as the level of automatic operation improves and with extensive full text data available, large-scale experiments in automatic indexing algorithms and techniques become more feasible.

The Parallel Processor (Pattern Articulation Unit)

The basic mission of the parallel processor is the rapid dissection and separation—the "articulation"—of image data into recognizable parts. It may be conceived, alternatively, as a programmable property filter which facilitates identification of morphological characteristics in two-dimensional data.

Physically, the PAU consists of a set of planar arrays of memory elements interconnected by extremely flexible, program-controlled logic circuitry. Each plane contains 1156 (= 34 × 34) bits of information which bear a one-to-one relation to the "geographical" coordinates of a picture plane (see Figure 2). However, the magnification factor between the original picture plane and the PAU plane is arbitrary—the area of a PAU plane may correspond, for example, to an entire document page at one extreme, to less than one square millimeter at the other extreme magnification.

Let us address a specific problem, a simple format analysis, to illustrate the application of the PAU. Suppose we wish to identify the geographical area occupied by a simple graphic sketch on a document page (see Figure 3a). Assume that any suitable sketch has less black (ink), on the average, per unit area than normal text.

The analysis which is simplified, to be sure, proceeds as follows:

(1) Insert a document page into the PAU with grayness value corresponding to the average grayness in the area which a memory cell represents.

(2) Mark with "X" all points having average or darker grayness. The marks appear in a separate plane as shown in Figure 3b, and cover (in general) all text areas. This operation is performed in parallel over the entire plane.

(3) Mark with "W" all areas which are not marked X, and which also extend from one edge directly (horizontally or vertically) to the opposite edge.

The operation marks W in all border zones and the vertical center strip (see Figure 3c).

The only remaining unmarked areas are relatively white, non-border areas, which correspond presumably to the desired sketch(es). If so, it is quite simple to

(4) Mark the corner points of the sketch region and read out the coordinates to the general purpose computer. The coordinates of the sketch boundary provide all the requisite information to (a) electronically snip out the sketch and reproduce it elsewhere, or perhaps (b) to introduce into thePAU the area immediately beneath the figure where its caption is probably located. The caption may be used for indexing.

The format analysis procedure indicated above could be accomplished in some 50 to 100 microseconds (exclusive of picture input time) with the present PAU design[7,8] which is by no means optimized for speed.

The other application of the PAU pertinent to the present discussion is that of character recognition. In this case, one character in black/white code is placed in a PAU plane where its morphological properties may be extracted and the object classified. The problems here center mostly on economics, the cost per character, especially over a great dispersion of type fonts, rather than on technical capability.

The Image Store

The critical attributes of an image store are large capacity and very low cost per stored information unit. Furthermore, the access time to any specified image must be as small as possible but certainly not in excess of about five seconds. Access time is inversely related to the maximum permissible number of simultaneous users, which, in turn, affects system economics. In any case, long optical tapes, i.e., serial stores, are hopelessly slow.

A semirandom access image store which satisfies the system design criteria is discussed next.

The basic storage unit is a film card consisting of a 300 mm. strip of 105 mm. sprocketed film. Stored microimages cover an area of about 85 × 250 mm. on the card with each image allotted a 1.5 mm. diameter field. Each image has about 1mm. maximum side length. Each card has a unique address corresponding to the shape of 12 slots along the upper card edge (see Figure 4). In the storage unit, the cards are suspended from semicircular selection rods which pass through the address coding holes of 4096 cards as shown in Figure 5.* The selection rods are analogous to the needle of the familiar needle-sort card addressing technique. Selection of one card is accomplished by simultaneous rotation of each of the 12 rods by 90°, either clockwise or counter-clockwise, which releases exactly one card. The released card falls due to gravity plus the frictional force of downward airflow which also serves to separate the cards. After falling a few centimeters, the selected card is engaged by a linear conveyor which whisks it into the reading station. The card's vertical position is monitored as it enters the read station at a speed of some six meters per second. As the addressed image row approaches the readout station, the card is stopped by a precision vernier "spear" mechanism which aligns the desired image row within ± 25 microns (approximately) of the readout station. The desired image is then illuminated through a fiber-optic pathway which projects the image through a fly's-eye magnification lens onto the face of a vidicon camera tube, where the image is converted to a high resolution television-picture signal.

Card replacement on the selection rods is accomplished by sprocket-driving the card upward from the read station into the replacement mechanism which then pushes the card onto the selection rods and returns to quiescent position. A storage capacity of 38 million images in one storage unit is well within present technical

---

*The selection mechanism is similar to that of National Cash Register Company's CRAM (Card Random Access Memory).[9]

feasibility of 1.5 mm. field diameter per image. Capacities up to 80 million images are practically possible.

The cycle time (access + replacement time) for this memory is estimated at 1 to 1.6 seconds for nearest and farthest card, respectively, from the readout station. Access from one image to another in the same row of the same card ("page-turning") is much faster, viz., about 120 milliseconds which is largely determined by the time required for the television camera to sweep a new image.

One Problem—Image Degradation

The manipulation of microimage cards in the image store raises the question of image degradation due to scratch, dust, etc. The meager experience available indicates that obscuration or deletion of small characters (e.g., period, comma) is frequently observed. One solution to this problem is the use of holograms rather than direct images. Every point of the direct image is then dispersed over the microimage area so that a defect only reduces the overall quality of the reconstructed image rather than obscuring any particular point. Holography entails some capricious problems in its own right, particularly with respect to mechanical alignment. For the immediate future, the system will employ a model direct image store while investigation of holographics matures.

The Dissemination Network and a Second Problem—Image Resolution

The design of the dissemination network has far-reaching critical ramifications, potentially, for the overall system design and economics. Inclusion of a wide-band network for dissemination of original image data has been considered only rarely—thus, the specific problems associated therewith have not been frequently discussed.

Image dissemination begins at the image store where a televised replica of a retrieved image is formed and transmitted over the network to a system user. It is assumed that each stored image is transmittable as one picture frame. The channel capacity or channel bandwidth affects the time required to transmit one picture frame with a specified resolution. The constraints are such that not every original document page can be transmitted as a single frame, from which it follows that one original document page may not, in general, be represented as one image in the image store.

Therefore, the questions which arise are (1) "How shall the network parameters be chosen to accommodate document images?" and

(2) "If some document images must be reorganized to 'fit' the network, how shall this be accomplished?"

Several pieces of data are required to answer the first question. Experiments to determine the speed and accuracy of televised character recognition by human subjects have been summarized by Shurtleff.[10] He finds that 98-99 percent correct symbol identification is attained for a vertical resolution of 8-12 lines per symbol height when the character visual size (to the observer) is 12 to 15 minutes of arc, and the horizontal resolution is relatively high. Figure 6 is a picture of text displayed at nine lines vertical resolution* for the tall characters "d," "p," "h," etc., and six lines for the short characters such as "s." In case the figure, as printed here, does not convey the original psychological impression, let me say that subjectively, the original display appears fully acceptable. We propose to allot 15 percent greater resolution than that shown by Figure 6, which results in approximately 10.5 vertical lines for tall lower case letters or 15 lines vertically† for a gross character line, i.e., including one-half the upper and lower line spacing.

The second necessary piece of data is the total resolution required adequately to reproduce text pages. Four cases exhibiting a range of difficulty are shown in Table 1.

On the basis of this data combined with technical state-of-the-art capability, we have chosen a dissemination network with 1536 lines vertical resolution, 1250 lines horizontal resolution transmitted at a bandwidth of 12 megahertz which allows one frame per 80 milliseconds. Crudely, pages of about the complication of a Time magazine page could be reproduced readably, whereas a page of Webster's Collegiate Dictionary at 13 lines per inch would not be reproducible as a single page. Very large original pages, pages with very small type or high resolution photographs would, in general, require segmentation or "reformatting." Before pursuing that problem, however, a few more remarks regarding the dissemination network are relevant.

The network described above is intended for relatively short distance transmission (e.g., up to two miles) where coaxial cable can be economically laid. If wireless transmission is used, it may be economically essential to operate at the standard§ commercial television (video) bandwidth of 4.2 megahertz, corresponding to one picture frame per 0.25 second with the same resolution as before. In any case, the network is not suitable for moving picture transmission without providing for switchable scanning speed at the camera and monitor. Compatibility with a national or international information interchange network is still an open question.

---

*Horizontal resolution was calculated at five lines for the width of the letter "s."

†All values are in scan lines and do not include a Kell factor.

§U. S. National Television Standards Committee.

TABLE 1
TELEVISION SYSTEM RESOLUTION REQUIRED FOR DIRECT REPRODUCTION OF EXAMPLE DOCUMENT PAGES

Document (Nominal Page)	Character Line/Inch	TV Linear Resolution/ Inch at 15 TV Lines/ Character Line	Total Page Dimension (Inches)		Total TV Linear Resolution Required for Reproduction		Reproducibility as a Single Frame
			Width	Height	Horizontal	Vertical	
Libraries of the Future-Licklider	5.5	82.5	5	8	410	660	Easy
1401 Programming McCracken	6	96	7.75	11	750	1060	Satisfactory
Time Magazine	7.5	115	8	11	920	1265	Satisfactory
Webster's Collegiate Dictionary 5th Edition	13	195	5.8	8.6	1100	1700	Impossible
		†			*		

*Assuming the aspect ratio (width/height) of picture is identical to the page.
†A resolution adequate for manual or mechanical recognition operations.

Before leaving the topic of television reproduction, we should also note Figure 7 which is a magnified reproduction of Figure 6 showing the approximate quality of characters which would be supplied for recognition processing.

Reformatting

Returning to the question of reformatting, we propose that any required reorganization of a page be performed prior to insertion in image store. Also, the reformatting would be controlled manually in the system's early phases with provisions for evolution toward fully automatic operation. All incoming documents would be accepted as COSATI-standard microfiche. Reformatting is performed, where necessary, and the reorganized images placed on second-level microfiche, after which batches of microfiche cards are reduced onto micro-image cards for the image store.

The reformatting operation in the mature stage is envisioned as follows. Original microfiche is fed through a scanner where each image, sequentially, is coarsely digitized under program control. The resulting numerical representation of the image is transferred to the parallel processor where format analysis proceeds along the lines described earlier. Average or maximum character heights may be determined by electronically "smearing" the image horizontally, which results in gray regions for print lines and white (or whiter) areas between lines. By counting the number of gray smears per inch in all regions of the image, the regions can be characterized according to their suitability for reproduction by the dissemination network.

The geographic data resulting from format analysis is applied to an optical microfiche camera and program-controlled iris which is capable of passing any desired rectangular area of the original microfiche onto secondary microfiche. The microfiche camera, therefore, records selected areas of the original microfiche at controlled magnification so that the secondary microfiche images are each transmittable through the dissemination network. Developed secondary microfiche cards are then transferred by a separate operation onto image store cards (see Figure 8). For all but the most pathological cases, one text page image need not be dissected into more than two images, which is a modest requirement. The reformatting operations are actually less difficult and certainly less time consuming per page, than character recognition. However, success in attaining a high level of automaticity depends upon the general purpose capability of the parallel processor and not simply upon a special purpose optical character reader.

The Remote Station

Cost is the principal ingredient in the design of any remote station. In the present context, the remote stations are optically passive—they originate no image data. Active communication from a remote station to the central facility is accomodated by a small keyset.

Visual data is displayed on a storage-type picture tube which will hold satisfactory resolution for one or two minutes. Hard-copy output will originate from a normal picture tube (non-storage type) slaved to the main display tube.

In order to copy a displayed page, the user will move a pointer to a spot on a surrogate microfiche card. This action moves the film holder of an adapted polaroid camera to a position where a standard microfiche image may be recorded. Then, upon activation of a "COPY" key, the image store address of the presently displayed page is reaccessed and scanned several times to reduce noise. Immediately thereafter, the shutter of the polaroid camera viewing the slaved picture tube is tripped. The microfiche card is then developed by polaroid† process, of course. It should also be noted that image positioning is at the user's discretion.

There are two modest problems concerning the hard-copy system as it is presently designed: (1) there is no microfiche card titling mechanism (other than purely manual), and (2) the resultant card size will not be COSATI-standard* because of limitations on readily available film sizes.

Conclusion

A general observation which our work suggests is that a high-speed dissemination system should not be regarded as a mere appendage to an information storage and retrieval system. The ramifications of the dissemination system characteristics reverberate throughout the system, particularly affecting the storage operation and the number of simultaneous users. These observations hold to some extent in any system employing block-organized data for dissemination.

The image-data oriented system is economically sensitive to dissemination system characteristics. If frame resolution is chosen very large, the transmission frame rate is reduced, adversely affecting the permissible number of simultaneous users. If frame resolution

---

†Using Polaroid 55 P/N film of 150 lines/mm. resolution.

*60 images per card rather than 72.

is low, the cost of reformatting images may be raised to an uneconomical level. In the present system, it is expected that less than 10 percent of the introduced data will require reformatting.

A means of format analysis, geometric object location and character recognition is essential for fully mechanized operation of an image-data oriented ISRD system. These data processing requirements can be satisfied by a programmable parallel processor (here, the Pattern Articulation Unit) using processing algorithms which are well-known in general but require specialization to text characteristics.

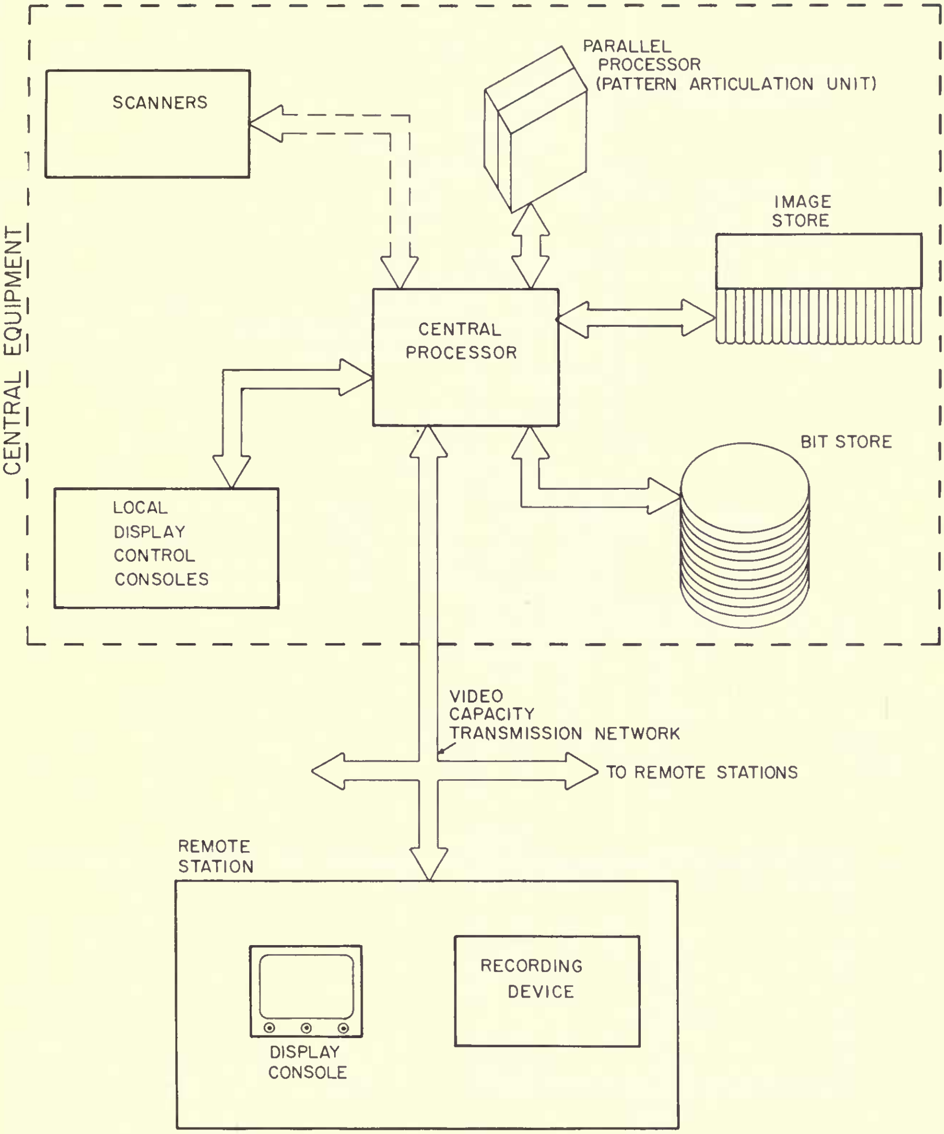


Figure 1
System configuration.

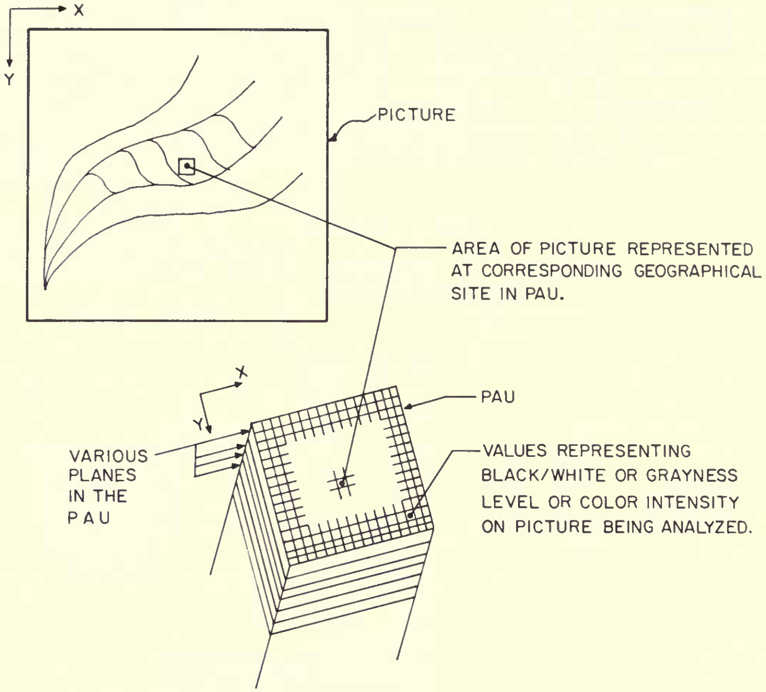
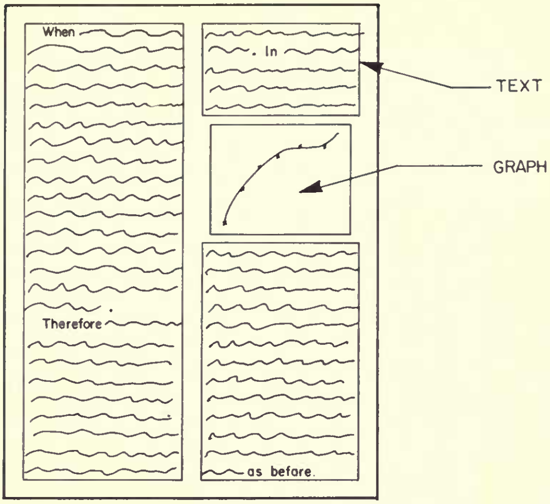


Figure 2
Parallel processor. Illustration of mapping a picture into PAU representation.



EXAMPLE OF A DOCUMENT
PAGE CONTAINING A
SIMPLE GRAPH SKETCH
FIG. 3a

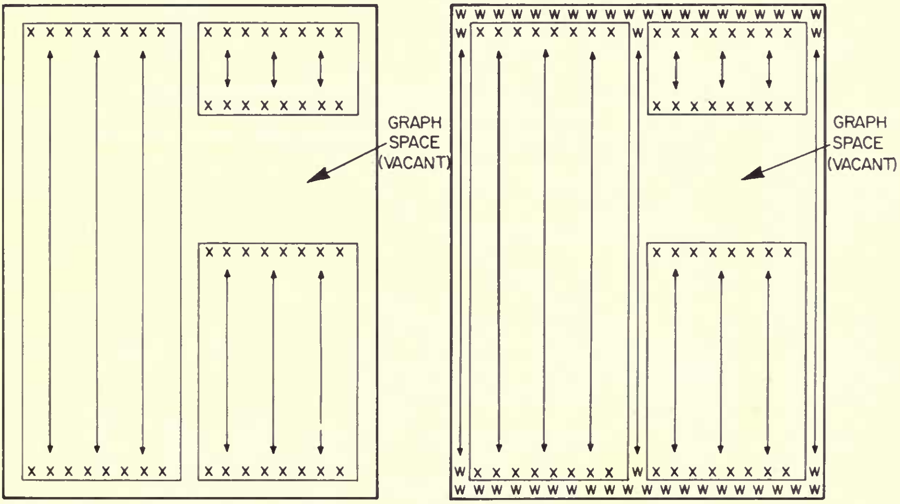


FIG. 3b
TEXT MARKED X

FIG. 3c
TEXT MARKED X
BORDER MARKED W

Figure 3
Format analysis

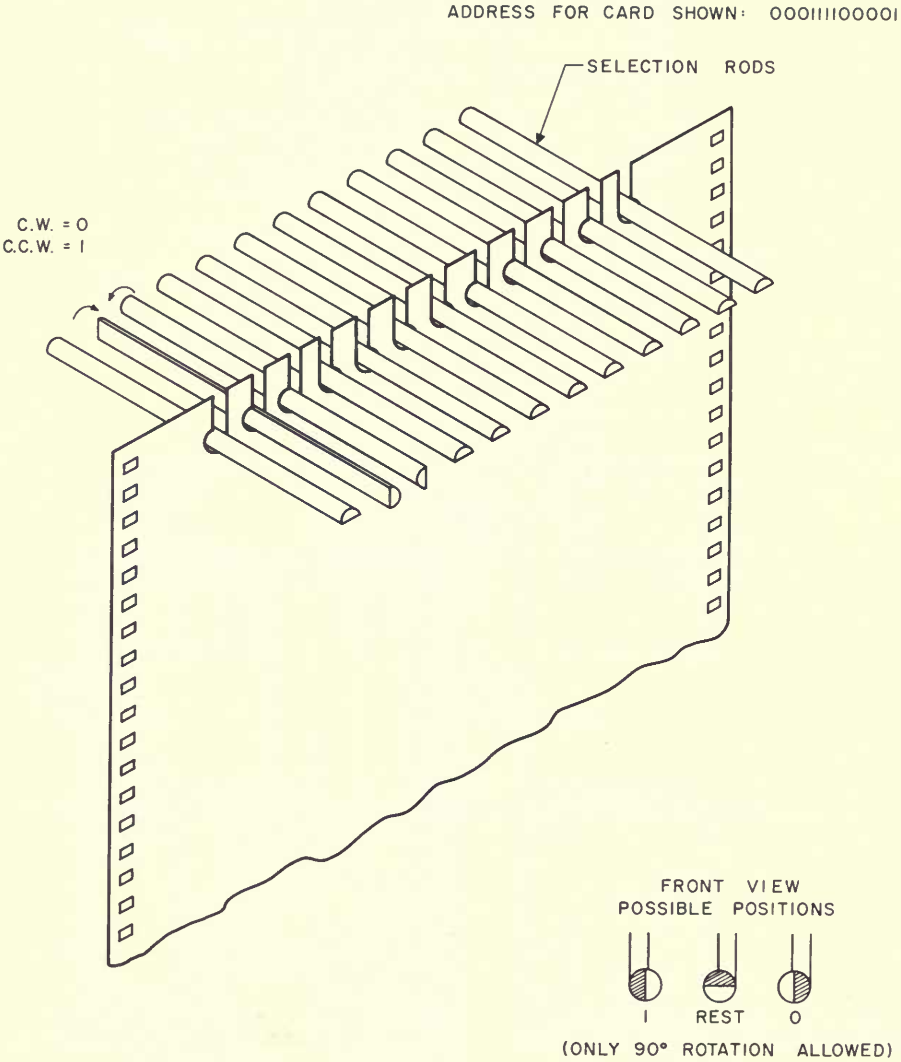


Figure 4
Design B: edge notched card

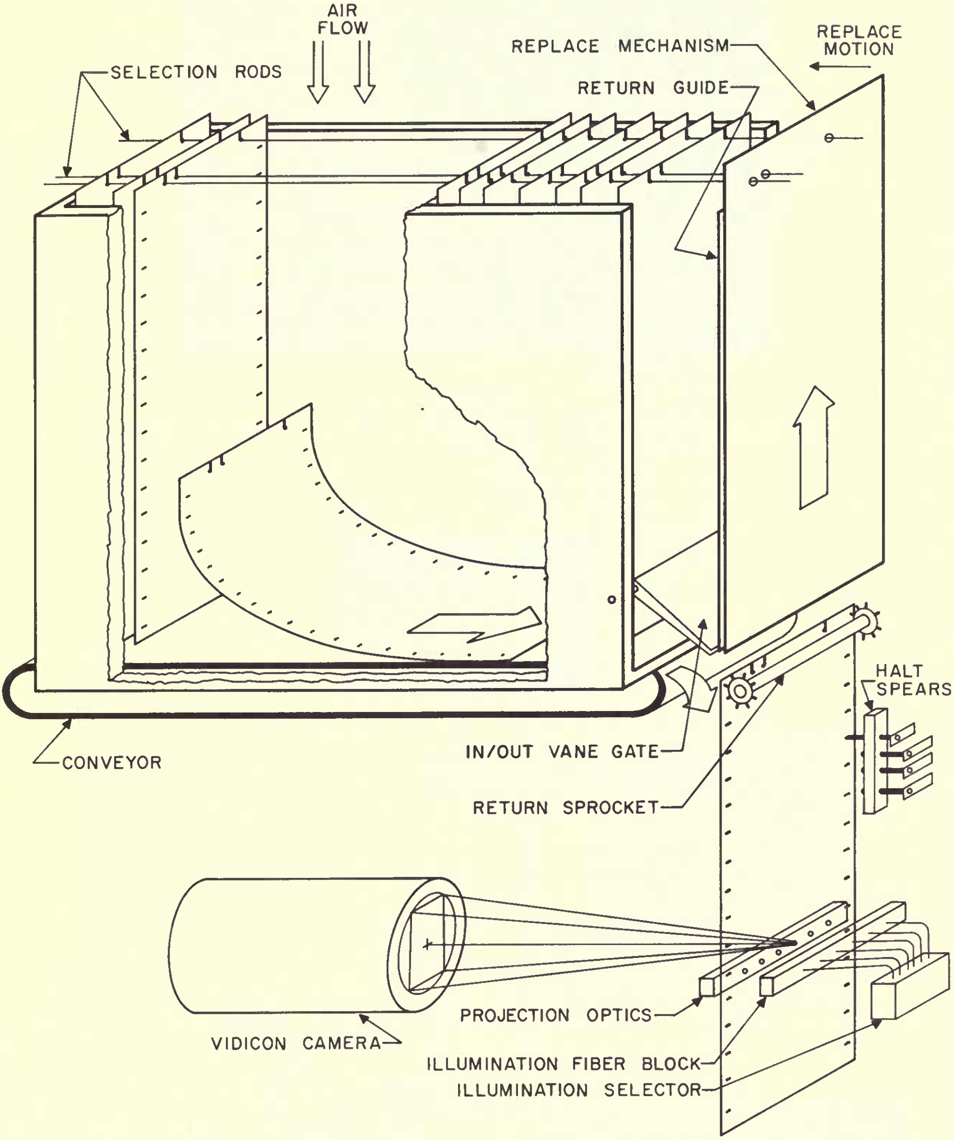
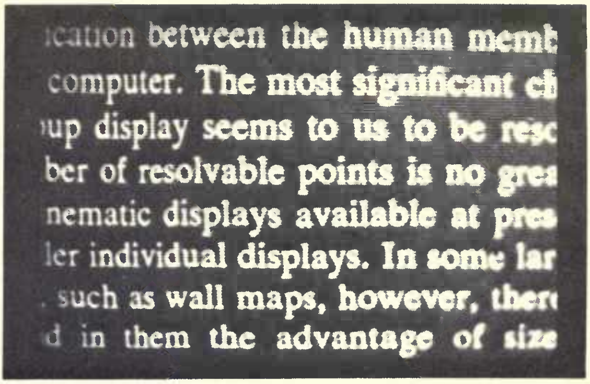


Figure 5
Design B: general mechanism location



ication between the human memb
computer. The most significant ch
up display seems to us to be resc
ber of resolvable points is no grea
nematic displays available at pres
ler individual displays. In some lar
. such as wall maps, however, there
d in them the advantage of size

Figure 6
TV monitor display of text: reproduced at 80% of proposed resolution.

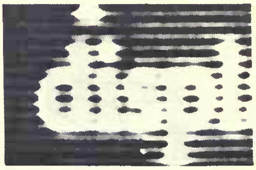


Figure 7
Enlarged view of fig. 6 illustrating character distortion

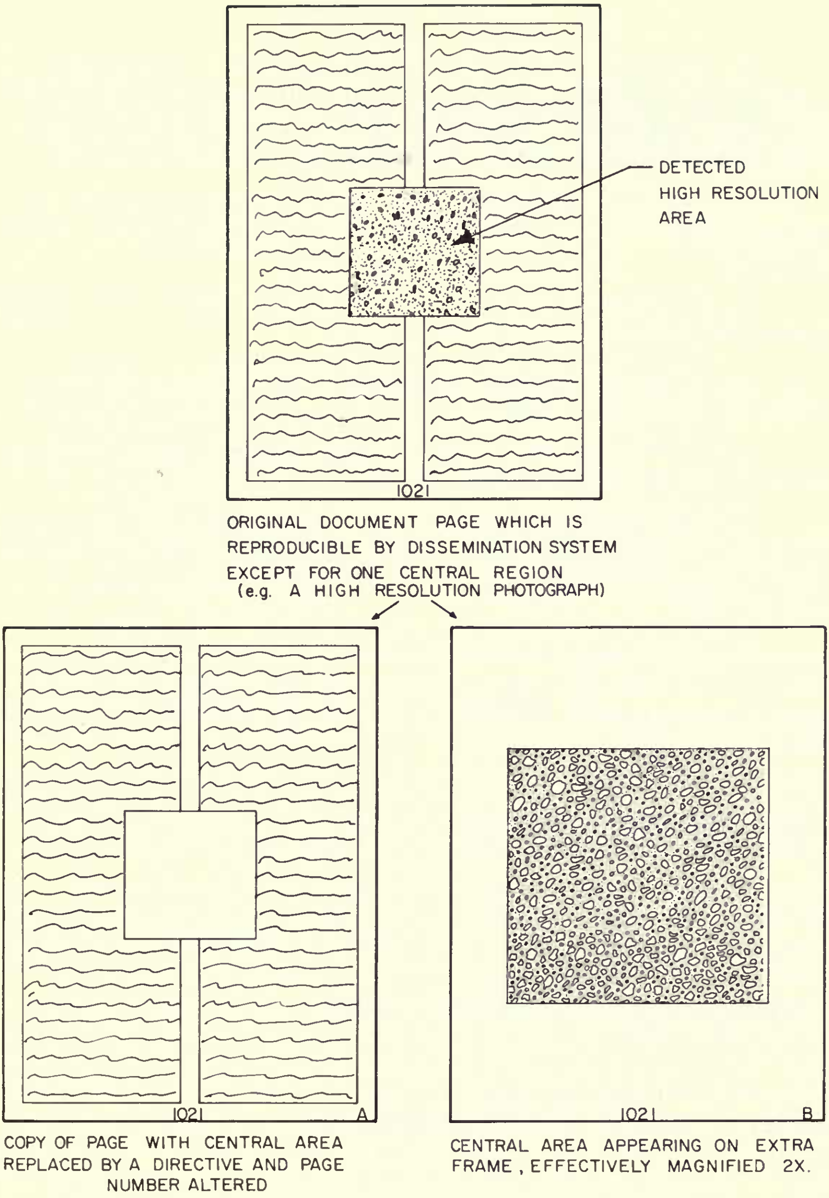


Figure 8
Reformatting example.

REFERENCES

1. McCormick, Bruce H., and Richardson, A. M. Design Concepts for an Information Resource Center with Option of an Attached Automated Laboratory. (Report No. 203, Department of Computer Science, University of Illinois.) Urbana, Ill., May 1, 1966.

2. Gelernter, Herbert. "Data Collection and Reduction for Nuclear Particle Trace Detectors," Advances in Computers, 6:229-296, 1965.

3. Overhage, Carl F. J., and Harman, R. Joyce, eds. INTREX. Cambridge, Mass., M.I.T. Press, 1965.

4. Licklider, J. C. R. Libraries of the Future. Cambridge, Mass., M.I.T. Press, 1965.

5. Ledley, Robert S. Use of Computers in Biology and Medicine. New York, McGraw-Hill, 1965.

6. Narasimhan, Rangaswamy. Syntactic Descriptions of Pictures and Gestalt Phenomena of Visual Perception. (Report No. 142, Digital Computer Laboratory, University of Illinois.) Urbana, Ill., July 25, 1963.

7. McCormick, Bruce H. "The Illinois Pattern Recognition Computer-ILLIAC III," Institute of Electrical & Electronics Engineers Transactions on Electronic Computers, EC-12:791-813, Dec. 1963.

8. Ray, Sylvian R., and Smith, K. C. Design of an Iterative Pattern Recognition Processor. (Report No. 161, Digital Computer Laboratory, University of Illinois.) Urbana, Ill., April 1964.

9. Bloom, Leon, et al., "Card Random Access Memory (CRAM): Functions and Use." In Proceedings of the Eastern Joint Computer Conference, Dec. 12-14, 1961, Washington, D. C. (American Federation of Information Processing Societies, Vol. 20: Computers—Key to Total Systems Control.) New York, Macmillan, pp. 147-157.

10. Shurtleff, Donald A. "Studies in Television Legibility—A Review of the Literature," Information Display, 4:40-45, Jan./Feb. 1967.